Aux origines des sciences cognitives



Aux origines des sciences cognitives

Sommaire

Jean-Pierre Dupuy

Aux origines des sciences cognitives



2005

Présentation

De la philosophie de l'esprit à l'intelligence artificielle, en passant par les neurosciences ou la nouvelle linguistique, les sciences cognitives forment aujourd'hui une nébuleuse de disciplines qui participent d'un renouvellement radical de nos façons de penser la connaissance et le savoir humain. Mais contrairement à ce qui est souvent dit, ce bouillonnement intellectuel n'est que secondairement lié à l'essor spectaculaire et récent de l'informatique. Il trouve son origine dans la cybernétique, née aux États-Unis dans les années quarante, au sein d'une petite communauté de scientifiques, neurobiologistes, mais aussi philosophes, psychanalystes et économistes.

Ce livre retrace cette histoire, jusque-là refoulée, qui conduit aux sciences contemporaines de l'esprit. On y découvrira en particulier la richesse des confrontations interdisciplinaires conduites au sein des fameuses « Conférences Macy », où discutèrent notamment Von Neumann, Wiener, McCulloch, Simon ou Rosenblueth.

Pour en savoir plus…

L'auteur

Jean-Pierre Dupuy, philosophe, est le directeur du Groupe de recherche et d'intervention sur la science et l'éthique (GRISE) de l'École polytechnique. Directeur de recherches au CNRS, professeur à l'Université Stanford (Californie) et maître de conférences à l'École polytechnique, Jean-Pierre Dupuy est l'auteur de nombreux ouvrages et notamment de La Panique (Les Empêcheurs de penser en rond, 1991), Le Sacrifice et l'Envie (Calmann-Lévy, 1992), Petite philosophie du tsumani (Seuil, 2005).

Du même auteur

La Panique, Les Empêcheurs de penser en rond, Paris, 1991.

Le Sacrifice et l'Envie, Calmann-Lévy, Paris, 1992.

Mécanisme mentaux, mécanismes sociaux, La Découverte, Paris, 1995.

Les Limites de la rationalité. Tome 1 : Rationalité, éthique et cognition (en codirection avec Pierre Livet), La Découverte, Paris, 1997.

Éthique et philosophie de l'action, Ellipses, Paris, 1999.

Introduction aux sciences sociales : logique des phénomènes collectifs, Ellipses, Paris, 1999.

Pour un catastrophisme éclairé : quand l'impossible est certain, Seuil, Paris, 2002, 2004.

Petite Métaphysique des tsunamis, Seuil, Paris, 2005.

La presse

Septembre 1956, les sciences cognitives voyaient le jour. Cette naissance n'avait rien d'une génération spontanée : pas moins de treize années ont été nécessaires. C'est l'histoire de cette gestation que Jean-Pierre Dupuy retrace. [...] L'histoire racontée par Jean-Pierre Dupuy est aussi faite de rendez-vous manqués avec la biologie, la phénoménologie, les sciences de l'homme et du social. L'analyse de ces échecs est aussi un appel lancé aux sciences cognitives pour qu'elles ne réitèrent pas les erreurs passées.

SCIENCES ET AVENIR

Collection

Ce livre a été précédemment publié, en 1994, dans la collection « Textes à l 'appui / Série sciences cognitives » aux Editions La Découverte.

ISBN numérique : 9782707172693
ISBN papier : 9782707147752

Ouvrage numérisé avec le soutien du Centre national du livre.

Ce livre numérique a été converti initialement au format EPUB le 07/03/2013 par Prismallia à partir de l'édition papier du même ouvrage.

S'informer
Si vous désirez être tenu régulièrement informé de nos parutions, il vous suffit de vous abonner gratuitement à notre lettre d'information bimensuelle par courriel à partir de notre site **http://www.editionsladecouverte.fr**, où vous trouverez l'ensemble de notre catalogue.



A la mémoire de Jean Ullmo
Pour Heinz von Foerster

Avant-propos

De 1946 à 1953, dix conférences, tenues à l'hôtel Beekman, 575 Park Avenue à New York, pour les neuf premières, à l'auberge Nassau de Princeton, New Jersey, pour la dernière, réunirent à intervalles réguliers certains des plus grands esprits de ce siècle. Organisées par la fondation philanthropique Josiah Macy Jr., elles sont entrées dans l'histoire sous le nom de conférences Macy. Mathématiciens, logiciens, ingénieurs, physiologistes et neurophysiologistes, psychologues, anthropologues, économistes, les membres de ce club fermé s'étaient donné pour ambition d'édifier une science générale du fonctionnement de l'esprit. Ce qui les avait amenés là, ce qu'ils ont pensé ensemble, ce qui est sorti de cette entreprise collective unique dans l'histoire des idées, tel est l'objet de ce livre.

Tout groupe de ce genre, pour affirmer son identité, doit se donner un nom de code. Dans le cas présent, ce fut « cybernétique ». Ce nom est aujourd'hui passé de mode, pour utiliser une litote. De 1954 au temps présent, le projet que s'était assigné le groupe cybernétique n'a cessé d'être remis sur le chantier sous des appellations chaque fois différentes. Depuis une quinzaine d'années, l'expression qui semble s'être imposée est celle de « sciences cognitives », ou parfois « sciences de la cognition » (*cognitive science* en anglais). Les raisons pour lesquelles les sciences cognitives, aujourd'hui, ont honte de leur ancêtre cybernétique constituent l'un des objets de mon enquête.

Mon intérêt pour le projet cybernétique date de 1976, année où j'eus la grande chance de rencontrer celui qui avait été, à partir de 1949, le secrétaire des conférences Macy, chargé d'en établir les comptes rendus : Heinz von Foerster. J'ai raconté ailleurs les circonstances de cette rencontre, qui fut décisive pour l'orientation de mes recherches[1] . A cette époque, von Foerster achevait une carrière tout entière consacrée à ce qu'il a appelé la seconde cybernétique, ou cybernétique du second ordre – carrière menée dans le cadre du centre de recherches qu'il avait établi à l'université de l'Illinois, à Urbana-Champaign : le Biological Computer Laboratory. C'est grâce à lui que j'entrai en contact avec deux biologistes dont les travaux sur l'auto-organisation du vivant avaient trouvé leur impulsion initiale dans ses propres idées : le Français Henri Atlan et le Chilien Francisco Varela. La vague des sciences cognitives n'avait pas encore atteint nos rivages et, pendant quelques années, ce fut dans notre pays la théorie des systèmes auto-organisateurs qui rassembla les énergies néo-ou post-cybernétiques. Un colloque, que j'organisai avec mon ami Paul Dumouchel, en juin 1981 au Centre culturel international de Cerisy-la-Salle, en Normandie, devait constituer l'un des points d'orgue de ce programme de recherches[2].

Sous l'impulsion de Jean Ullmo, l'École polytechnique avait décidé de se doter d'un centre de recherches philosophiques, avec une forte composante épistémologique : ce fut, en 1982, la création du CRÉA, le Centre de recherche en épistémologie appliquée[3], avec une équipe choisie parmi ceux qui avaient contribué à l'organisation et au succès du colloque de Cerisy. Jean Ullmo, qui fut mon maître à l'École, avait déjà beaucoup donné à celle-ci, par ses recherches et par son enseignement. C'est en particulier grâce à sa détermination que la vénérable institution avait consenti à introduire la science

économique dans le cursus de ses élèves. La véritable passion de Jean Ullmo était cependant la philosophie des sciences, comme en témoigne son livre, devenu un classique, *La Pensée scientifique moderne.* Son décès, en 1981, ne lui aura malheureusement pas permis de présider aux premiers pas du CRÉA qui, je l'espère, auraient comblé ses vœux. C'est tout naturellement que je dédie ce livre à sa mémoire. C'est non moins naturellement que je le dédie aussi à Heinz von Foerster, dont l'enthousiasme contagieux fut pour beaucoup dans les choix théoriques que je fus amené à faire au moment de la constitution du CRÉA.

Les théories de l'auto-organisation étaient donc présentes dans les premières recherches du centre. Mais bientôt, ce fut la vogue des sciences cognitives. Les responsables de la recherche de notre pays comprirent, bien qu'un peu tard, l'importance de ce programme de recherches interdisciplinaires qui bousculait les frontières établies de longue date, amenant au dialogue et à la confrontation les neurophysiologistes et les psychologues, les philosophes et les ingénieurs, les anthropologues et les linguistes. J'eus la chance de rencontrer alors un logicien, Daniel Andler, dont la quête philosophique s'était cristallisée sur les sciences cognitives. Andler amena au CRÉA un groupe de chercheurs éminents qui avaient fait le même choix que lui. C'est ainsi que notre centre est devenu l'un des tout premiers lieux de recherche, en France, en sciences cognitives, avec une forte composante philosophique. L'originalité du CRÉA, dans notre pays et même sans doute dans le monde, tient à ce que s'y côtoient et s'y confrontent des démarches et des paradigmes très variés. La philosophie analytique, dans sa composante philosophie de l'esprit, y joue un rôle important, mais cela n'exclut pas la phénoménologie. La réflexion philosophique ne craint pas de s'y nourrir de modèles mathématiques – à l'instar de la première cybernétique. Pour des raisons de circonstances et de jeux de pouvoir, l'institutionnalisation des sciences cognitives s'est faite, en France, autour de la neurobiologie et de l'intelligence artificielle, reléguant à une seconde place les sciences de l'homme et de la société. Là encore, le CRÉA occupe une place singulière, par l'importance attribuée à la philosophie sociale et politique, ainsi qu'aux sciences sociales proprement dites, essentiellement l'économie et l'anthropologie. Celles-ci, dès les conférences Macy, voient leur sort lié à ce mécanicisme de notre temps.

C'est dans ce contexte que j'éprouvai le besoin de réfléchir aux origines de ce mouvement d'idées dont mes choix, mais aussi les circonstances, m'avaient fait l'un des artisans. L'occasion m'en fut donnée par Dominique Wolton qui dirigeait au CNRS le programme Science-Technologie-Société (S-T-S). Celui-ci me confia en 1983 la responsabilité d'un programme de recherche généalogique sur les théories de l'auto-organisation. Je constituai aussitôt une équipe composée d'Isabelle Stengers, de Pierre Livet et de Pierre Lévy. Les deux Pierre devaient exploiter le très riche matériau que représentaient les travaux du Biological Computer Laboratory (BCL), cette Mecque de la seconde cybernétique ; Isabelle, forte de sa collaboration avec Ilya Prigogine, avec qui elle venait de publier *La Nouvelle Alliance*, devait quant à elle retracer l'histoire de la notion d'auto-organisation dans les sciences physico-chimiques. Je me réservai l'analyse de contenu des Actes publiés des conférences Macy, auxquels j'avais pu avoir accès grâce à la générosité de Heinz von Foerster.

La dynamique de la recherche collective allait complexifier ce schéma initial. C'est en novembre 1985 que nous remîmes notre rapport, sous la forme de deux numéros spéciaux des *Cahiers du CRÉA.* Il comportait deux importants textes de synthèse rédigés par

Isabelle Stengers et par Pierre Livet ; cinq précieuses monographies signées par Pierre Lévy sur divers aspects des travaux du B C L, mais aussi sur la vie et l'œuvre de Warren McCulloch, ce neuropsychiatre qui fut le véritable artisan des conférences Macy et, donc, de la première cybernétique ; et enfin ma propre contribution, intitulée « L'essor de la première cybernétique (1943-1953) ».

Quand Heinz von Foerster sut que je me lançais dans ce travail, il me prodigua ses plus vifs encouragements car, me dit-il, en dépit de leur très grande importance historique, les conférences Macy n'ont fait jusqu'ici l'objet d'aucune étude systématique. La situation est en train de changer rapidement, ajouta-t-il cependant – nous étions en 1983 – car un historien des sciences américain, Steve Heims, a décidé de consacrer ses forces et son temps à une reconstitution aussi fidèle que possible des faits, et de leurs circonstances, qui marquèrent l'histoire de cette période de pionniers. Je pris donc contact avec Heims et nous nous vîmes une première fois à Boston en 1984. Steve Joshua Heims était un physicien d'origine allemande qui s'était installé aux États-Unis après avoir fui le nazisme. Il avait rompu avec la physique pour des raisons déontologiques, jugeant que l'orientation générale des recherches faisait de cette discipline un simple moyen au service des techniques de puissance et de mort. Il avait donc décidé de prendre du recul et de faire de l'histoire des sciences. Son choix s'était porté sur deux personnalités scientifiques hors pair qui avaient marqué profondément la science, la technique et la société de leur temps, avant, pendant, et après la Seconde Guerre mondiale : John von Neumann et Norbert Wiener. Quand je rencontrai Heims, il venait de publier un ouvrage remarqué se présentant comme la biographie parallèle des deux grands mathématiciens, significativement intitulé : *John von Neumann and Norbert Wiener. From Mathematies to the Technologies of Life and Death*[4]. Mais, surtout, cette première recherche lui avait donné envie de faire un point complet sur ce qu'avait été le projet de la cybernétique, puisque l'un et l'autre de ses deux personnages avaient joué un rôle crucial dans la création de cette discipline. Il avait donc décidé de consacrer un nouvel ouvrage aux conférences Macy.

Pendant toute la période où je travaillais sur mon rapport, je restai en étroit contact avec Heims. En 1985, je l'invitai à présenter un cycle de conférences au CRÉA. J'incluai dans notre rapport final un texte de lui qui se présentait comme le résumé de son livre à venir et s'intitulait « An Encounter between Neo-Mechanists and the Human Sciences »[5]. Sans le travail d'historien de Heims, je n'aurais pu mener à bien ma propre recherche.

Les années passèrent, surtout marquées en France par le développement vigoureux des sciences cognitives et l'ouverture du CRÉA à ce nouveau champ, comme je l'ai expliqué ci-dessus. En 1991, Steve Heims publiait enfin son ouvrage attendu sur les conférences Macy, sous le titre : *The Cybernetics Group*[6]. C'est ce moment que choisirent mon ami et collègue du CRÉA Jean-Michel Besnier et François Gèze, directeur des éditions La Découverte, pour me demander de reprendre et de compléter mon rapport de 1985 aux fins d'une publication.

La question se posait évidemment de l'utilité d'un second ouvrage sur les conférences Macy. Après avoir longtemps délibéré, j'optai pour une réponse positive, et ce pour au moins deux raisons.

La première raison, c'est qu'en dépit de ma grande admiration pour le travail patient et obstiné d'historien qu'a accompli Steve Heims, je suis en désaccord avec l'option fondamentale qu'il a choisie. Heims plonge la cybernétique dans le contexte socio-

économique, politique et idéologique de l'Amérique d'après guerre – ce qui en soi est intéressant – et décide de n'y voir qu'un reflet de ce contexte – ce qui est beaucoup plus problématique. Je crois trop au pouvoir des idées et à leur dynamique autonome pour me satisfaire en général de la perspective « externaliste » en sociologie des sciences, mais dans le cas présent, cette perspective me paraît introduire un très sérieux biais. Pour des raisons de circonstances que j'expose dans le cours de ce livre en tirant parti des informations précieuses rapportées par Heims, la cybernétique naissante a dû s'allier à un mouvement d'idées, ou plutôt un véritable lobby à l'américaine, qui prétendait assurer la paix mondiale et la santé mentale généralisée au moyen d'un bizarre cocktail fait de psychanalyse, d'anthropologie culturelle, de physique de pointe et des idées nouvelles que la cybernétique apportait ! Heims se déchaîne contre cette idéologie scientiste, naïve et individualiste qu'il juge très peu… *politically correct.* Le problème est qu'il consacre l'essentiel de son ouvrage aux membres du « groupe cybernétique » qui appartenaient à ce mouvement. De là un décentrement qui obscurcit ce qu'était le projet proprement scientifique et philosophique de la cybernétique. Il ne s'agissait de rien de moins que de conduire l'aventure scientifique à son apothéose en édifiant une science de l'esprit. C'est bien cette ambition qui fait de la cybernétique l'ancêtre des sciences cognitives. Si cette ambition a pu s'épanouir à ce moment précis de l'histoire, c'est beaucoup moins, à mon sens, en raison du contexte social, politique et idéologique du moment que comme conséquence d'une évolution longue qui porte sur la représentation que se fait l'Occident de l'activité de connaître. C'est du moins la thèse que je défends dans le premier chapitre de ce livre. La cristallisation s'est produite à ce moment-là parce qu'il a fallu le choc des grandes découvertes logiques des années trente pour mener le processus à maturation.

J'avais éprouvé un malaise semblable à la lecture du premier livre de Steve Heims où il oppose un méchant von Neumann à un bon Wiener (je caricature quelque peu) : tandis que le premier fait monter la température de la « guerre froide » en travaillant à mettre au point la bombe à hydrogène, le second conçoit des prothèses pour mal-entendants… La question des rapports circulaires entre science, technique et société a certes de l'importance, mais je ne crois pas qu'elle doive occulter l'analyse de la dynamique endogène des mouvements d'idées.

La seconde raison découle immédiatement de la première. Heims est si occupé à dénoncer le capitalisme et son projet universel de maîtrise sur les êtres et les choses qu'il en néglige, ou en tout cas fait passer au second plan, la question que j'ai placée au cœur de mon enquête : dans quelle mesure la cybernétique peut-elle être tenue pour la matrice dont sont issues ce qu'on appelle aujourd'hui les sciences cognitives ? Cette question, on le verra, est controversée, et je tente d'y apporter une réponse en présentant les prolégomènes à ce que serait une histoire intellectuelle des sciences cognitives. Cela m'amène fréquemment à sauter en dehors de la période qui fait directement l'objet de mon enquête pour la confronter à des protagonistes ou à des développements ultérieurs. J'espère avoir évité la plupart des pièges qui guettent tout exercice d'histoire rétrospective, et je crois en tout cas ne pas avoir trop cédé à la facilité qui consiste à présenter le précurseur comme en état de manque par rapport au savoir qui surgira plus tard. Ce qui m'en a gardé est peut-être le sentiment qu'il y avait dans la phase cybernétique une richesse de débats et d'intuitions auxquels les sciences cognitives d'aujourd'hui feraient parfois bien de s'alimenter. Ce que je reproche à la cybernétique, ce n'est pas de ne pas avoir su ce que l'on découvrirait bien après elle, c'est de ne pas

s'être saisie de savoirs la concernant qui étaient à sa portée, à condition qu'elle veuille bien s'apercevoir de leur existence.

Mon projet répond donc à des soucis très différents de ceux de Steve Heims, et c'est pourquoi j'ose penser que ce livre a sa place à côté du sien. Les réserves que j'ai exprimées à son sujet ne m'empêchent pas de redire ma dette : sans les matériaux patiemment rassemblés par Heims sur de nombreuses années, mon enquête eût perdu beaucoup de ses fondements.

Outre Steve Heims, que tous ceux auxquels j'ai fait référence, nominativement ou non, au cours de cet historique de mon travail, reçoivent ici l'expression de ma gratitude : mes collègues du CRÉA qui m'ont beaucoup apporté par leurs recherches en sciences cognitives et par les échanges que j'ai eu la chance d'avoir avec eux : Daniel Andler, Paul Dumouchel, Pascal Engel, Françoise Fogelman, Pierre Jacob, Maurice Milgram, Jean Petitot, Joëlle Proust, François Recanati, Dan Sperber et Gérard Weisbuch ; ceux qui m'ont fait confiance en commanditant mon travail et en faisant preuve de beaucoup de patience à mon égard : Jean-Michel Besnier, François Gèze et Dominique Wolton ; mes compagnons de recherche lors de la première phase de mon travail, à qui j'ai emprunté certaines de leurs réflexions ou découvertes : Isabelle Stengers, Pierre Lévy et Pierre Livet ; et enfin, ceux qui ont su m'initier à l'esprit de la cybernétique : Henri Atlan, Francisco Varela et Heinz von Foerster. C'est grâce à la générosité de la DRET et du Dr Guy Véron que j'ai pu mener les recherches qui m'ont permis de transformer le rapport de 1985 dans le présent livre. L'aide éditoriale de Laurence Helleu m'a été précieuse.

Ces remerciements n'impliquent pas que je sois resté fidèle à ce que les uns ou les autres m'ont appris, ni que je partage leurs options théoriques ou philosophiques. A la fin de son ouvrage, Steve Heims me range parmi les *happy few* qui aujourd'hui portent le flambeau allumé par la première cybernétique. C'est me faire un redoutable honneur, que je ne suis pas sûr de mériter. Mon jugement, on le verra, n'est pas spécialement tendre sur l'histoire que je raconte. Il m'arrive de contester les témoignages, ou de critiquer les prises de position, de certains de ceux qu'ici je remercie. Ce livre se présente finalement comme l'histoire d'un échec. Échec grandiose, si l'on veut, ou en tout cas productif et riche d'enseignements – mais échec néanmoins, si l'on compare ce qui fut réalisé aux ambitions de départ. Je disais ci-dessus que j'avais pris garde d'éviter les pièges de l'histoire rétrospective. C'est plutôt un nouveau genre, assez hasardeux, que j'ai pratiqué : l'histoire contractuelle. La science cognitive naissante aura raté tellement de rendez-vous importants que la tentation est grande de se demander fréquemment : « Que se serait-il passé si seulement… ? » Exercice gratuit, peut-être, trop facile, assurément, mais qui traduit la frustration de l'interprète.

Stanford, mai 1994

Les documents suivants, auxquels il sera souvent fait référence, seront désignés par les symboles que voici :

Macy 6 : *Cybernetics – Circular Causal and Feedback Mechanisms in Biological and Social Systems.* Transactions of the Sixth Conference, March 24-25, 1949, New York. Edited by Heinz von FOERSTER. Josiah Macy, Jr. Foundation, New York, 1950. 209 p.

Macy 7 : *Cybernetics – Circular Causal and Feedback Mechanisms in Biological and Social Systems,* Transactions of the Seventh Conference, March 23-24, 1950, New York. Edited by Heinz von FOERSTER, Margaret MEAD and Hans Lukas TEUBER. Josiah Macy, Jr. Foundation, New York, 1951. 251 p.

Macy 8 : *Cybernetics – Circular Causal and Feedback Mechanisms in Biological and Social Systems,* Transactions of the Eighth Conference, March 15-16, 1951, New York. Edited by Heinz von FOERSTER, Margaret MEAD and Hans Lukas TEUBER. Josiah Macy, Jr. Foundation, New York, 1952. 240 p.

Macy 9 : *Cybernetics – Circular Causal and Feedback Mechanisms in Biological and Social Systems,* Transactions of the Ninth Conference, March 20-21, 1952, New York. Edited by Heinz von FOERSTER, Margaret MEAD and Hans Lukas TEUBER. Josiah Macy, Jr. Foundation, New York, 1953. 184 p.

Macy 10 : *Cybernetics – Circular Causal and Feedback Mechanisms in Biological and Social Systems,* Transactions of the Tenth Conference, April 22, 23 and 24, 1953, Princeton. Edited by Heinz von FOERSTER, Margaret MEAD and Hans Lukas TEUBER. Josiah Macy, Jr. Foundation, New York, 1955. 100 p.

Tel. Mech. : *Teleological Mechanisms,* Conference held by the New York Academy of Sciences, October 21-22, 1946. *Annals of the New York Academy of Sciences,* vol. 50, art. 4, p. 187-278. October 13, 1948.

Hixon : *Cerebral Mechanisms in Behavior.* The Hixon Symposium (California Institute of Technology, September 1948). Edited by Lloyd A. JEFFRESS, John Wiley and Sons, New York, 1951.

McC. Intro : Introductions aux œuvres complètes de Warren MCCULLOCH. Intersystems Publications, Salinas (Californie), 1989. (Cette édition est malheureusement aujourd'hui épuisée, l'éditeur ayant disparu. Je l'ai consultée dans les archives personnelles de Heinz von Foerster).

1

La fascination pour le modèle

« Geometrica demonstramus quia facimus ; si physica demonstrare possemus, faceremus. »

Gianbattista VICO, *De nostri temporis studiorum ratione.*

Herbert Simon, Prix Nobel d'économie, est généralement considéré comme l'un des pères fondateurs de l'intelligence artificielle. Les économistes utilisent son concept de « *rationalité limitée* », les théoriciens de la décision et les professeurs de gestion ont fait leur celui de *satisficing*, les spécialistes des sciences cognitives connaissent bien le programme d'ordinateur qu'il a conçu, avec Alan Newell, pour simuler les processus créatifs de la pensée humaine lorsqu'elle se saisit d'un problème en termes de fins et de moyens – programme qu'il a éloquemment dénommé le *General Problem Solver.* Les historiens des idées le tiennent pour un homme de la Renaissance égaré au XXe siècle, une sorte de Leonardo da Vinci des temps actuels. Le titre d'un de ses principaux ouvrages, *The Sciences of the Artificial,* publié en 1969, trahit assez bien son ambition intellectuelle.

De nombreux autres titres jalonnent sa bibliographie, qui sont construits selon un schéma identique : *Models of Man* (1957), *Models of Discovery* (1977), *Models of Thought* (1979, 1989), *Models of Bounded Rationality* (1982). On ne s'étonne pas que livrant récemment au public son autobiographie, il l'ait intitulée : *Models of my Life* (1991). Curieux livre en vérité. S'y révèle un homme qui a construit sa vie en plein accord avec les principes qui guidaient sa pensée – une pensée qui affirme d'elle-même que son meilleur *modèle* est l'ordinateur digital. Ainsi, comme tout savant de son importance, Herbert Simon a beaucoup voyagé. Il affirme cependant qu'il n'a jamais rien appris lors de ses séjours à l'étranger qu'il n'eût pu apprendre, ou n'ait déjà appris, plus vite, plus aisément et à moindre coût dans les livres de n'importe quelle bonne bibliothèque des États-Unis. Cela, qui correspond à son expérience, est vrai pour tout Américain adulte moyen, énonce-t-il sentencieusement sous la forme du « théorème des voyages[1] ». Nous apprenons que la première fois que le couple Simon visita l'Europe, en 1965, déjà bien imprégnés de la vérité de ce théorème, ils s'arrangèrent pour ne rien voir qu'ils ne connaissaient déjà par les livres ou les tableaux. Les deux semaines qu'ils passèrent en France furent une sorte de pèlerinage dédié à Proust et à Cézanne. La Sainte-Victoire, l'Estaque : ils ne se rendaient que dans les endroits exacts, « au mètre près », où Cézanne avait planté son chevalet : « Nous n'apprîmes rien de nouveau ; nous avions déjà vu les tableaux. » Et Simon de citer Oscar Wilde : « Où étaient les brouillards de Londres avant que Turner les peignît ? », pour conclure : « La Nature, à son habitude, imite l'art. » Dès la première page de son livre, il nous avoue sa fascination pour Jorge Luis Borges, ce virtuose des labyrinthes et des jeux de miroirs. Borges, l'auteur du « Thème du traître et

du héros », pour qui l'histoire n'est qu'une copie de la littérature.

La vertu des modèles

De bonnes anthologies d'intelligence artificielle font de Thomas Hobbes le précurseur de la discipline et consacrent leur premier chapitre à un extrait du *Léviathan*. L'imputation est loin d'être absurde ou anachronique. Hobbes conçoit l'État ou *Commonwealth* sur le modèle d'un « automate qui se meut par ressorts et rouages comme une horloge ». Qu'est-ce que le Léviathan ? Un immense « homme (ou animal) artificiel ». Les sciences cognitives se reconnaissent volontiers dans la formule : « *Thinking is reckoning* » (Penser, c'est calculer). Or derrière cette communauté de mots, il y a une véritable parenté philosophique.

Le système de Hobbes est déjà habité de l'idée que Vico formulera dans les termes célèbres : « *Verum et factum convertuntur* » (« Ce qui est vrai et ce que l'on fait sont convertibles »). Nous ne pouvons connaître rationnellement que ce dont nous sommes la cause, que ce que nous avons fabriqué. A l'origine, le principe du *verum factum* s'entend sur le mode du manque : nous ne pourrons jamais connaître la nature comme Dieu, car celui-ci l'a créée et nous ne pouvons que l'observer. Bientôt, cependant, le principe acquiert une valeur positive, plus en conformité avec l'affirmation croissante du subjectivisme moderne. Ce que l'homme fait, il peut le connaître rationnellement, de façon démonstrative et déductive, malgré la finitude de son entendement. Par ordre décroissant de perfection de la connaissance, les mathématiques viennent, selon ce critère, en premier, cela ne fait pas de doute, suivies cependant non par les sciences de la nature, mais par les sciences morales et politiques. C'est ainsi que Hannah Arendt, dans des pages célèbres, interprète aussi bien la conception artificialiste de la politique développée par Hobbes que le passage opéré par Vico « de la science naturelle à l'Histoire, seule et unique sphère, pensait-il, où l'homme pourrait obtenir la connaissance certaine puisqu'il n'y aurait affaire qu'aux produits de l'activité humaine[2] ».

Selon Arendt, cependant, la science de la nature elle-même devait être dès les commencements orientée par la conviction qu'on ne peut connaître qu'en faisant, ou plutôt qu'en re-faisant. « Dès le début […] le savant aborda la nature du point de vue de Celui qui l'a créée[3] . » L'insistance sur le comment des processus plutôt que sur l'être des choses s'explique ainsi, mais aussi et surtout le rôle considérable dévolu à l'expérimentation par la science. « Pour utiliser l'expérimentation afin de connaître, il fallait déjà être convaincu que l'on ne peut connaître que ce que l'on a fait, car cette conviction signifiait que l'on peut s'informer des choses que l'homme n'a point faites en *se représentant* et en *imitant* les processus qui les ont amenées à l'existence[4]. »

Se représenter, imiter, répéter, reproduire : c'est bien ce que fait l'expérience scientifique, on en conviendra. Il est cependant étonnant que Hannah Arendt n'ait pas vu qu'il y a un *faire* dans l'activité scientifique, bien plus universel que celui que l'on trouve dans l'expérimentation et qui confirme à merveille ses analyses : je veux dire la fabrication de *modèles*.

La science est la seule activité humaine où le mot « modèle » a le sens inverse de celui que lui donne la langue usuelle. Est modèle ce que l'on imite, ou ce qui mérite d'être imité. Or le modèle scientifique est au départ une imitation. Il a avec la réalité le même

type de rapport que celui qu'entretient un « modèle réduit » avec l'objet dont il est la reproduction plus aisément manipulable. L'inversion de sens est frappante et mérite d'être méditée. On dispute beaucoup en philosophie et histoire des sciences de la nature et des fondements de la connaissance scientifique. Que la science comme *activité* consiste essentiellement à se construire des objets sous la forme de modèles est en revanche une vérité incontestable, bien que trop peu connue des non-scientifiques. Qu'est-ce qu'un modèle ? Ne jamais partir des définitions, enseigne Karl Popper, et on serait bien en peine d'en donner une ici qui ne soit pas balayée par le mouvement incessant de la *création* scientifique. Disons simplement qu'il s'agit d'une idéalité, le plus souvent formalisée et mathématisée, synthétisant un système de relations entre des « éléments dont l'identité et même la nature est, jusqu'à un certain point, indifférente, et qui peuvent, par suite, être changés, remplacés par d'autres éléments analogues ou différents, sans que (le modèle) soit altéré[5] ».

Le modèle est comme une forme abstraite qui vient à s'incarner ou à se réaliser dans les phénomènes. Des domaines très différents de la réalité phénoménale, comme l'hydrodynamique et l'électricité, la lumière et les vibrations sonores, peuvent se représenter par des modèles identiques, ce qui établit entre eux une relation d'équivalence. Le modèle est la classe d'équivalence correspondante. Cela lui donne une position de surplomb, telle une Idée platonicienne dont le réel n'est que la pâle copie. Mais le modèle en science, c'est l'homme qui le fait. Voilà comment l'enchevêtrement de l'imitant et de l'imité se réalise. Le modèle scientifique est une imitation humaine de la nature que le savant prend bientôt pour « modèle » – au sens ordinaire – de celle-ci.

Si, comme le veut Arendt, le scientifique est un *Homo faber*, c'est avant tout parce qu'il est concepteur et fabricant de modèles. Cela lui confère une maîtrise à laquelle il n'oserait prétendre sur les phénomènes eux-mêmes. Maîtrise explicative et prédictive, tout d'abord, grâce à la puissance de l'outil mathématique. Celui-ci donne à l'exploration des propriétés du modèle – exploration qui prend la forme d'expériences de pensée – une efficacité, une rapidité et une élégance que n'atteindra jamais l'expérimentation faite sur le monde phénoménal. Maîtrise du pouvoir créateur de l'analogie, surtout. Étant donné deux domaines dont on postule qu'ils sont représentables par le même modèle, ce que l'on sait de l'un suggère de livrer l'autre à des expériences nouvelles, permet de formuler à son sujet des hypothèses inédites et de découvrir des propriétés intéressantes. Ainsi, « lorsque Bohr eut été amené à l'hypothèse d'une similitude entre la structure du noyau formé de nucléons (protons et neutrons) et celle d'une goutte d'eau formée de molécules, les lois connues de l'évaporation se traduisirent en celles de la désintégration radioactive, les conditions où la gouttelette se sépare en deux firent prévoir celles de la fission ; la bombe atomique devait sortir de ces travaux[6] ».

Le modèle *abstrait* de la réalité phénoménale le système des relations fonctionnelles qu'il juge seules pertinentes, mettant pour ainsi dire entre parenthèses tout ce qui ne relève pas de ce système et en particulier, nous l'avons vu, le nombre, l'identité et la nature des éléments qui sont en relation. C'est ainsi que le même modèle peut représenter la structure du système solaire ou celle de l'atome de Rutherford. Le rôle que jouent les diverses formes de *fonctionnalisme* en science (y compris en sciences cognitives) a sa source dans la pratique universelle de la modélisation. On voit aussi que le modèle, par constitution, est susceptible de réalisations matérielles multiples. Inversement, un même phénomène peut faire l'objet de modélisations sans rapport l'une avec l'autre, bien que

parfaitement équivalentes quant aux résultats des calculs auxquelles elles mènent. L’un des exemples les plus célèbres dans l’histoire de la science est sans doute l’étude du mouvement d’un point matériel dans un champ de forces. Le système d’équations différentielles de la mécanique newtonienne est un premier modèle, le formalisme lagrangien ou hamiltonien un second. Celui-là postule des forces à distance s’exerçant dans un espace vide et passif ; celui-ci suppose au contraire un espace actif, un « champ » qui exerce son influence sur les objets qui y sont plongés. On sait que le formalisme du second modèle a pu, par une généralisation assez immédiate mais, au départ, non vraiment réfléchie, engendrer la mécanique quantique, exploit que le premier modèle n’eût pu permettre de réaliser[7].

Les modèles ont donc une vie propre, une dynamique autonome découplée de la réalité phénoménale. Dans son activité modélisatrice, le savant projette son esprit dans le monde des choses. Les énormes succès de la modélisation scientifique sont comme un témoignage que l’esprit est à la fois distinct de la matière, et adéquat à elle.

Cela ne va pas sans danger. Le modèle est tellement plus pur, tellement mieux maîtrisable que le monde des phénomènes : le risque existe qu’il devienne l’objet exclusif de l’attention du savant. Des théories, voire des disciplines entières peuvent s’organiser autour de l’étude des propriétés d’un modèle. Ainsi la science économique, qui n’en finit pas d’explorer les ressources, certes impressionnantes, du modèle de l’équilibre général que Walras lui a légué. Il n’est pas certain que la nouvelle-née « science de l’esprit » échappe à ce travers.

Manipuler des représentations

Venons-en à l’esprit, précisément. Après les mathématiques, les sciences morales et politiques (y compris l’histoire), les sciences de la nature et de la vie, le *verum factum* ne pouvait manquer d’affecter la connaissance de l’esprit, lorsque la science se jugea en mesure de s’attaquer à ce bastion ultime de la Création. Ce moment fut celui où elle se crut enfin capable d’apporter une réponse acceptable par elle à ce vieux problème qu’elle avait abandonné à la philosophie : celui des rapports entre l’âme et le corps ou, pour le dire dans le vocabulaire d’aujourd’hui, entre la matière et l’esprit.

L’esprit n’est pas une production de l’esprit, la conscience ne s’est pas faite elle-même. C’est pourquoi Vico reconnaissait en fait une dimension d’opacité irréductible dans le *cogito* et même, par voie de conséquence, dans le domaine social-historique[8] . L’ironie de l’histoire des idées est grande. Pour que naisse la conviction contraire que l’esprit humain peut se connaître lui-même rationnellement, parce qu’il peut concevoir et fabriquer une réplique de lui-même, il aura fallu que des découvertes logiques mettent à mal la validité du *verum factum* dans le seul domaine où elle n’avait jamais été mise en doute : les mathématiques. Jusqu’aux travaux des logiciens des années trente, on pouvait répéter ce que Hobbes avait écrit dans le *De homine* : « C’est parce que nous créons nous-mêmes les figures que la géométrie se trouve considérée comme de l’ordre du démontrable[9]. » Après les travaux de Gödel et de Turing, il faut renoncer à cette évidence. La théorie hobbienne de la définition génétique ne s’applique pas systématiquement aux mathématiques. Nous sommes la cause des êtres mathématiques et, cependant, certains d’entre eux maintiennent par rapport à nous une obscurité irréductible. Or c’est de cette découverte prodigieuse, qui

n'a pas fini de produire ses effets dans l'histoire de la pensée, que date le point de départ de la nouvelle science de l'esprit. Nous allons y revenir.

Il est bien connu que le paradigme classique en sciences cognitives s'est développé autour de la « métaphore de l'ordinateur ». C'est le genre d'antienne que l'on répète beaucoup trop, car cette métaphore n'était nullement nécessaire et, ce livre va le montrer, les choses ont en fait commencé avant que l'ordinateur existe – ou, plus précisément, alors qu'il existait en tant qu'objet matériel technique, mais qu'on ne disposait pas encore d'une théorie fonctionnaliste de cet objet. Cette théorie qui nous est devenue si familière, par laquelle nous distinguons le « logiciel » (*software*) du « matériel » (*hardware*), est un produit de la révolution conceptuelle qui marque l'avènement des sciences cognitives, et non sa source.

Connaître, c'est produire un modèle du phénomène et effectuer sur lui des manipulations réglées. Toute connaissance est reproduction, représentation, répétition, simulation. Cela, nous l'avons vu, caractérise le mode scientifique, rationnel de connaissance. Les sciences cognitives font de ce mode le mode unique de toute connaissance. Pour elles, tout « système cognitif » se rapporte au monde comme le savant à son objet. On ne s'étonne pas que la notion de représentation occupe une place centrale dans la science de la cognition. L'analogie, cependant, est encore plus profonde. Soit un système cognitif matériel : savant, homme, animal, organisme, organe, machine. Ce qui fait que ce système connaît par modèles et représentations doit pouvoir lui-même être modélisé en abstrayant du substrat matériel chaque fois différent le système de relations fonctionnelles responsable de la faculté de connaître. Le fonctionnalisme de la science de la cognition se situe donc à (au moins) deux niveaux logiquement emboîtés : celui de la représentation élémentaire ; et celui de la représentation de la faculté de représentation. C'est à ce second niveau qu'une science de la cognition peut tout à la fois se déclarer matérialiste ou physicaliste et revendiquer son autonomie par rapport aux sciences de la nature (et de la vie). L'esprit, compris comme le modèle de la faculté de modéliser, a retrouvé sa place dans l'univers matériel. Ou, pour le dire en d'autres termes, aujourd'hui plus familiers, il y a de l'information (et même du sens) dans le monde physique. Les facultés de l'esprit ne sont jamais que les propriétés de systèmes à traiter de l'information.

Connaître, c'est effectuer, sur des représentations, des manipulations réglées. Cette proposition est fidèle à l'esprit du paradigme dominant en sciences cognitives et cependant, il lui manque l'essentiel. L'essentiel, c'est la nature *logique* des manipulations et des règles en question. Le modèle scientifique, avons-nous dit, prend le plus souvent une forme mathématique – et, plus précisément encore, il se ramène à un système d'équations différentielles reliant des grandeurs. L'époque qui précède immédiatement l'histoire que nous allons conter a produit des modèles mathématiques tant du système nerveux que des circuits électriques. Il aura cependant fallu le génie de McCulloch et Pitts d'un côté, de Shannon de l'autre, pour comprendre que la modélisation pertinente était en fait de type logique – donc que l'on pouvait décrire en termes logiques le fonctionnement de certains systèmes matériels, mais qu'inversement ces systèmes matériels pouvaient se représenter comme réalisant, voire incarnant la logique, cette forme supérieure de la pensée.

La machine de Turing

Au commencement était le Logos, certes, mais celui-ci était une machine. Lorsque Alan Turing publie en 1936 son fameux article « On Computable Numbers, with an Application to the *Entschei-dungssproblem*[10] », il n'a certes pas conscience d'énoncer les prolégomènes d'une nouvelle science de l'esprit. Son objectif est de résoudre un problème de logique posé par Hilbert, celui de la « décision » (*Entscheidung*) : étant donné une formule quelconque du calcul des prédicats, existe-t-il un procédé systématique, général, effectif permettant de déterminer si cette formule est démontrable ou non ? Après 1931, date de la publication du non moins fameux article de Kurt Gödel, « Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme[11] », la question acquiert une actualité nouvelle. Gödel, en effet, y établit le théorème dit « d'incomplétude » : pour tout système formel au sens de la logique (ce qui implique la donnée d'un langage formel, de formules de ce langage tenues pour axiomes et de règles d'inférence) suffisamment riche pour contenir l'arithmétique, on a la propriété suivante : ou le système est inconsistant (il conduit à des théorèmes contradictoires), ou il existe au moins une proposition vraie qui n'est pas démontrable dans le système. La formule qui lui correspond n'est donc pas un théorème, pas plus, évidemment, que sa négation, à laquelle correspond une proposition fausse. On aura beau « dérouler » l'ensemble des théorèmes du système, on ne trouvera « jamais » ni la formule ni sa négation. En restant à « l'intérieur » du système, on ne sera « jamais » fixé sur la valeur de vérité de la formule, laquelle est donc « indécidable ».

Pour établir ce résultat, Gödel a accompli un geste décisif. Il a montré qu'il était possible de coder par des nombres entiers les formules, mais aussi les séquences de formules, comme celles qui constituent des démonstrations. La proposition qui affirme qu'une formule donnée est démontrable peut ainsi s'exprimer sous la forme d'une proposition arithmétique. Cette arithmétisation de la logique donne un fondement rigoureux à la maxime : raisonner, c'est calculer (sur des nombres entiers). Cependant, le théorème d'incomplétude laisse en suspens la solution au problème posé par Hilbert. L'appréciation de la portée du théorème dépend en effet crucialement de ce que l'on entend par « calcul effectif », « procédure finie », etc., puisque ce que le théorème affirme, c'est l'*inexistence* de telles procédures opératoires, constructives, décidables pour répondre à des questions sur les systèmes formels, à l'intérieur même de ceux-ci. En d'autres termes, il manque encore une définition rigoureuse, mathématique, de l'*algorithme*. Or sans une telle définition, il n'y a pas de réponse définie au problème de Hilbert.

En 1936, on dispose de plusieurs caractérisations de ce qu'est une fonction calculable de façon effective, dont une proposée par Gödel lui-même : la « récursivité générale », la « lambda-définissabilité », etc. Le résultat remarquable est que ces diverses caractérisations, qui sont d'esprit très différent, définissent la même classe de fonctions. Elles sont équivalentes, au sens qu'elles sont coextensives. Il est donc tentant de postuler que cette classe unique correspond à la notion intuitive de calculabilité effective. Cette année-là, Alonso Church franchit le pas, indépendamment de Turing[12]. Bien évidemment, l'équivalence, à elle seule, ne prouve rien. Il se pourrait que chaque caractérisation, par coïncidence, manque la cible de la même façon – la cible, c'est-à-dire la formalisation de la notion intuitive d'une procédure automatique, soumise à des règles fixes, totalement maîtrisable par la raison, n'impliquant ni sens, ni interprétation, ni créativité. En vérité, il n'y a rien à prouver. Entre des définitions mathématiques et une notion préthéorique,

l'adéquation n'est pas de l'ordre du démontrable. C'est pourquoi la suggestion de Church selon laquelle toute fonction calculable (au sens intuitif) est une fonction récursive (au sens mathématique) est présentée comme une *thèse* et ne peut être autre chose.

Il manque à ce bouillonnement d'idées un élément décisif – décisif, en particulier, pour l'histoire des sciences cognitives qui nous occupe. On s'étonne rétrospectivement qu'il n'ait pas paru plus évident à l'époque. Cette notion de calculabilité effective que l'on chercher à cerner, où n'entre que de l'exécution aveugle et « automatique », c'est manifestement le fonctionnement d'une *machine* qui en donne l'image la plus claire. C'est à Turing que nous devons d'avoir pris cette métaphore mécanique au sérieux. Dans sa thèse remarquable sur « La constitution de la théorie des automates », Jean Mosconi fait une conjecture intéressante sur la nature de la résistance qu'une telle idée rencontrait encore dans les années trente : « Eu égard aux calculatrices existant à l'époque – une fois oubliées les perspectives ouvertes un siècle plus tôt par Babbage – la référence aux possibilités de calcul d'une machine pouvait apparaître comme un rétrécissement arbitraire de l'idée de calculabilité [...] Si pour nous le sens naturel de "calculabilité mécanique" est "calculabilité par une machine", il est vraisemblable que, jusqu'à Turing, "mécanique" était employé plutôt métaphoriquement et ne voulait rien dire de plus que "servile" (épithète qui qualifie encore couramment l'exécution d'un algorithme)[13] » Dix ans plus tard, dans la période qui va nous occuper, ces freins antimécanistes auront totalement sauté.

Dans son article de 1936, Turing propose une formalisation mathématique de la notion abstraite de machine. Suivant la tradition, on appellera le modèle formel ainsi défini « machine de Turing ». Nous allons dire dans un instant ce que fait et ce que peut faire une machine de Turing. Mais Turing lui-même est persuadé qu'il a, avec ce concept, dégagé l'essence de la procédure mécanique. La thèse qu'il énonce est que toute procédure mécanique est modélisable par machine de Turing ; ou, puisque ces procédures sont « fonctionnelles » (aux deux sens du mot : l'un qui renvoie au fonctionnalisme précédemment défini, l'autre à la fonction mathématique) et qu'il s'agit en définitive de calcul : toute fonction calculable mécaniquement est calculable par une machine de Turing. Cette « thèse de Turing », pas plus que celle de Church, ne peut évidemment être démontrée. La calculabilité mécanique est une notion vague, intuitive, tandis que la « Turing-calculabilité » est définie formellement.

Ce qui peut être démontré en revanche, et l'est par Turing un an plus tard[14], est le résultat remarquable suivant : la classe des fonctions calculables par machine de Turing coïncide avec la classe des fonctions récursives, elle-même équivalente à celle des fonctions lambda-définissables. La dernière-née des caractérisations formelles de la notion de calculabilité effective se révèle équivalente à toutes les autres. Ce qui, tout à la fois, renforce la thèse de Turing, corrobore celle de Church, rend vraisemblable que ces deux thèses n'en font qu'une et, surtout, révèle la profonde parenté qui lie la notion de calculabilité effective à celle de procédure mécanique, celle-ci, étant la plus intuitive, fournissant la clé de celle-là.

Turing s'est donné les moyens de résoudre le problème de la décision de Hilbert. Il lui suffit de traduire : « Existe-t-il une procédure générale effective permettant de déterminer si une formule du calcul des prédicats est démontrable ou non ? » par « Existe-t-il une machine (de Turing) capable de décider ce qu'il en est ? » Cette nouvelle formulation revient à traiter le système formel au sens de la logique, c'est-à-dire un ensemble

d'axiomes et de règles d'inférence, comme une procédure mécanique produisant de nouvelles formules appelées théorèmes.

Il convient à ce stade de rappeler ce qu'est une machine de Turing. Suivant le cas ou l'humeur, on en trouvera la description compliquée ; ou bien, au contraire, on s'émerveillera qu'un être aussi primitif puisse prétendre à représenter de façon synthétique tout ce que nous entendons intuitivement par procédure finie, effective ou décidable. Une machine de Turing comprend trois organes : la machine *stricto sensu*, susceptible à tout moment (le temps est discrétisé) de se trouver dans un état, dit « état interne », appartenant à une liste finie ; un ruban illimité dans les deux sens, représentant la mémoire de la machine : ce ruban est divisé en cases, chacune d'entre elles comportant ou non une certaine marque ; et, enfin, une tête de lecture-écriture-effacement capable de réaliser les opérations suivantes : placée devant l'une des cases, elle lit si celle-ci contient la marque ou non ; elle peut effacer la marque si elle existe, ou l'inscrire dans le cas contraire ; elle peut se déplacer d'une case vers la droite ou vers la gauche.

Le fonctionnement de la machine est régi par un tableau invariant. En entrée, nous y trouvons la « configuration » de la machine, c'est-à-dire le couple constitué à un moment donné de l'état interne et du contenu de la case lue (présence ou absence de la marque). En sortie, complètement déterminés par la configuration d'entrée, nous avons les trois mouvements effectués par la machine qui la mènent au moment suivant : son changement (éventuel) d'état interne ; le maintien ou le changement du contenu de la case lue ; le déplacement (ou l'absence de déplacement) de la tête. Enfin, certaines configurations d'entrée donnent à la machine l'instruction de s'arrêter.

Moyennant codage, une machine de Turing donnée (par son tableau) réalise une fonction numérique particulière. L'argument entier de la fonction est représenté par les marques qui figurent sur le ruban lorsque l'on lance la machine ; celle-ci, étape après étape, actualise, suivant ses instructions, l'état du ruban ; lorsque la machine s'arrête (si c'est le cas), celui-ci donne, après décodage, la valeur de la fonction pour l'argument de départ. La machine de Turing est donc une calculatrice arithmétique. Grâce au travail de Gödel, elle est par là même beaucoup plus que cela. Puisque la logique est arithmétisable, la machine de Turing est aussi un calculateur symbolique. Elle constitue le modèle de la pensée symbolique.

On aura noté que la machine de Turing n'est pas un objet fini : son ruban, c'est-à-dire sa mémoire, est illimité. On peut s'en étonner, car l'objectif visé est précisément de modéliser une procédure effective, donc finie. Comme l'exprime très bien, ici encore, Jean Mosconi, « les machines de Turing sont surprenantes à deux égards : en ce que des structures aussi sommaires puissent faire tant de choses ; en ce que des machines dites "infinies" ne puissent pas "tout faire". En fait, les machines de Turing ne sont que des machines finies pourvues d'une mémoire infinie. Si on injecte trop d'infini, on détruit les notions mêmes de machine et de calcul, qui doivent toujours garder quelque chose de finitiste. Ce serait le cas par exemple si on admettait qu'une machine puisse exécuter une infinité d'opérations en un temps fini. On peut dire que la machine de Turing est intéressante parce qu'elle est, en quelque sorte, raisonnablement infinie[15] ». A quoi il faut ajouter que si l'on dotait la machine d'une mémoire simplement finie, ses performances deviendraient beaucoup moins « intéressantes ». Nous aurons l'occasion d'y revenir.

Revenons au problème de la décision. Turing montre qu'il est équivalent, dans la formulation mécaniste qu'il propose, au problème de « l'arrêt de la machine de Turing » :

soit une machine de Turing et un état initial du ruban, existe-t-il une procédure mécanique – donc, « thèse de Turing » oblige, une autre machine de Turing – capable de déterminer si la première machine va s'arrêter, ou, au contraire, boucler indéfiniment ? On veut donc construire une machine qui, recevant pour entrée sur son ruban la donnée en question, réponde « oui » (moyennant codage) si et seulement si la première machine s'arrête pour cette donnée. Turing fournit la démonstration qu'une telle machine n'existe pas. Il répond ainsi par la négative à la question posée par Hilbert.

Cette découverte est à l'égal de celles que nous avons déjà rencontrées au long de notre parcours : inattendue, déconcertante. Soit l'ensemble des productions d'une machine de Turing (l'ensemble des valeurs de la fonction numérique dont elle est le modèle pour tous les arguments qui la font s'arrêter). Cet ensemble est, par définition, engendré mécaniquement (on dit qu'il est « récursivement énumérable »). Et, cependant, il n'est pas caractérisable, globalement, de façon mécanique (à la question : « Telle valeur fait-elle partie de l'ensemble ? », il n'y a pas de réponse mécanique). En termes imagés : ce que la machine est capable d'engendrer est au-delà du mécanique. Les meilleurs esprits du temps furent bouleversés par cette découverte. Nous y reviendrons.

Tous ces résultats eurent un impact considérable – non, certes, immédiatement sur la culture au sens large, ni même sur la communauté scientifique ou philosophique dans son ensemble. Le petit nombre, s'il brûle d'une foi totale pour un projet théorique, suffit à déplacer les montagnes de l'histoire des idées. C'est ainsi que, dans les années quarante, naquit la cybernétique – et, dans son sillage, les sciences cognitives.

Ces *happy few* étaient au départ une poignée de mathématiciens, d'ingénieurs, de physiologistes. Certains d'entre eux furent, dans leur domaine, des génies. Mais le projet qu'ils avaient en commun, et qui donnait sens et vigueur à leurs diverses entreprises, était primordialement philosophique.

Avec le recul des ans, il est tentant de dire que ce projet consistait à donner une réponse scientifique, donc matérialiste, au vieux problème philosophique des rapports entre l'âme et le corps. Il y a beaucoup de vrai dans cette interprétation, mais il faut prendre garde aux illusions rétrospectives. C'est seulement beaucoup plus tard que la révolution conceptuelle introduite par la machine de Turing servit de base à une solution fonctionnaliste au problème des rapports entre la matière et la pensée. S'appuyant sur les travaux des maîtres de l'intelligence artificielle, tels Herbert Simon et Alan Newell, défendu et promu par des « philosophes de l'esprit » aussi exaltés que Jerry Fodor et Zenon Pylyshyn, ce « fonctionnalisme computationnel-représentationnel » revient à prendre au pied de la lettre la « métaphore de l'ordinateur » – l'ordinateur, cette machine de Turing incarnée dans des circuits électroniques. L'esprit, ou plutôt chacune de ses facultés particulières (par exemple : la capacité de former le concept général de triangle et de subsumer des occurrences triangulaires particulières sous ce concept), est conçu comme une machine de Turing opérant sur les formules d'un langage interne, privé, analogue à un langage formel en logique. Les symboles (ce qui est écrit sur le ruban de la machine) ont un triple mode d'existence. Ils sont matériels (incarnés dans la neurophysiologie) et donc soumis aux lois de la physique (d'abord soumis à celles de la neurophysiologie, mais celle-ci est supposée pouvoir être réduite à celle-là) ; ils ont une forme et, en tant que tels, ils sont régis par des règles syntactiques (analogues aux règles d'inférence dans un système formel au sens de la logique) ; ils sont enfin dotés de sens et ont une valeur sémantique. L'abîme qui semble séparer le monde physique du monde des

significations est franchi grâce au niveau intermédiaire que constitue la syntaxe, c'est-à-dire le monde des processus mécaniques, celui précisément dans lequel l'abstraction qu'est la machine de Turing est opérante. Le parallélisme entre processus physiques soumis à des lois causales et processus mécaniques réalisant des opérations de calcul, ou syntactiques, ou inférentielles, ne fait pas mystère dès lors qu'on se convainc qu'il existe dans le monde matériel des machines de Turing incarnées : l'ordinateur, certes, mais aussi tout processus naturel pouvant être considéré comme récursif[16] . Le parallélisme entre syntaxe et sémantique est garanti quant à lui par les théorèmes logiques de consistance et de *complétude* : les processus mécaniques qui réalisent les règles syntactiques maintiennent la cohérence interne des représentations symboliques, ainsi que leur adéquation à ce qu'elles représentent. Maintien ne veut évidemment pas dire création, et l'une des pierres d'achoppement de ce modèle informatique est la question de savoir comment le sens vient aux symboles. Le fonctionnalisme turingien constitue le cœur de ce qu'on appelle le « cognitivisme », qui est encore aujourd'hui le paradigme dominant dans les sciences de la cognition. Il est l'objet de nombreuses et profondes critiques, les plus sévères provenant parfois de ceux qui ont été parmi ses fondateurs[17].

Mais à l'époque que nous considérons dans ce livre, celle des pionniers, c'est surtout par ce qu'elle dit des rapports entre la pensée et la machine, plutôt que des rapports entre la pensée et la matière, que la thèse de Turing transporte les esprits. Telle est du moins la lecture que nous allons défendre et illustrer. La pensée, cette activité psychique, cette faculté de l'esprit ayant la connaissance pour objet, n'est rien d'autre finalement qu'un processus mécanique réglé, un automatisme « aveugle » – faut-il ajouter « bête » ? S'agit-il de dévaluer l'homme ? D'élever la machine ? Ou, au contraire, de faire de l'homme un démiurge capable de créer un cerveau ou un esprit artificiel ? Chacune de ces interprétations a sans doute sa part de vérité, plus ou moins grande selon les individus et les époques.

Les zélateurs, les idéologues vont donner de la thèse de Turing des versions de plus en plus fortes, exprimées dans des termes de plus en plus imprécis. On la présentera comme quelque chose de *démontré*. « Des découvertes logiques prouvent qu'il n'y a rien d'inconcevable dans l'idée d'une machine (ou de matière) pensante » : ce genre d'affirmation, littéralement fausse ou privée de sens, puisque, encore une fois, on n'est pas ici dans l'ordre du démontrable, se retrouve jusque sous la plume des meilleurs. On dira aussi : tout ce que l'esprit humain peut accomplir et qui peut être décrit avec précision, sans ambiguïté, en un nombre fini de mots, est exécutable par un ordinateur convenablement programmé ; ou encore (en 1965, sous la plume de Herbert Simon) : « En l'espace de vingt ans, les machines seront capables d'effectuer toutes les tâches, quelles qu'elles soient, que peuvent effectuer les hommes. »

Il serait facile d'ironiser. Les années ont infligé un démenti cinglant à cette déclaration présomptueuse et vaine. Ne soyons pas trop sévères. Pour accomplir ce qui l'a été, et qui n'est pas négligeable, il fallait ce que Karl Popper a appelé un « programme métaphysique de recherche », sorte d'acte de foi qui fournit principes, cadre de pensée et surtout impulsion à l'investigation scientifique. La thèse de Turing, en dépit, ou plutôt à cause des dérives idéologiques auxquelles elle se prête aisément, a su rallier suffisamment d'énergies et d'intelligences pour que naisse une science matérialiste et mécaniste de l'esprit.

Les découvertes logiques au sens strict auraient pu donner lieu à un tout autre type de

lecture, et les meilleurs esprits du temps, tel John von Neumann, l'ont bien compris. Si les *théorèmes* de Turing montrent quelque chose, c'est bien que penser n'est pas nécessairement calculer. Soit la fonction numérique qui, pour une machine de Turing donnée, prend la valeur 1 pour tout argument tel que la machine s'arrête, 0 dans le cas contraire. On peut *penser* cet être mathématique, en former clairement et distinctement le concept ; et, cependant, insolvabilité du problème de l'arrêt de la machine de Turing oblige, il échappe au calcul. On ne peut donc le *connaître*, si on entend par connaissance, conformément au principe du *verum factum*, l'acte consistant à se mouler dans le principe générateur de l'objet. Le modèle mécanique de Turing eût pu servir à approfondir la distinction fondamentale entre penser et connaître, propre à la philosophie critique. On peut regretter qu'il ait été trop souvent le prétexte d'un affrontement stérile entre ceux pour qui le slogan « penser, c'est calculer » fait office de signe de ralliement, et ceux qu'il fait s'étrangler d'indignation.

Connaître, c'est-à-dire simuler

Dans ce même article de 1936, Turing démontre l'existence d'une machine de Turing singulière (il en existe en vérité une infinité) qu'il nomme « universelle » : elle est en effet capable d'imiter, de mimer, de reproduire, de *simuler* le comportement de n'importe quelle autre machine de Turing. Il suffit d'inscrire sur son ruban un symbole (ensemble de marques) qui code le tableau de fonctionnement de la machine à imiter (coder par un nombre ou une séquence de marques – un ensemble de règles ou d'instructions est une opération concevable et réalisable effectivement, on le sait grâce à Gödel) : pour tout arguaient (variable d'entrée) inscrit sur son ruban, à côté du code de la machine à imiter, la machine de Turing universelle produira le même résultat que celle-ci.

Chaque machine de Turing particulière imite une faculté particulière de l'esprit : elle en est le modèle. La machine de Turing universelle imite n'importe quelle machine de Turing particulière : elle constitue le modèle des modèles. Nous avons appris plus haut que le fonctionnalisme de la science de la cognition définit l'esprit comme le modèle de la faculté de modéliser. La machine de Turing universelle est donc le modèle de l'esprit. Ou doit-on dire qu'elle *est* esprit ?

Plutôt que « penser, c'est calculer », la formule qui résume l'esprit des sciences cognitives devrait être : « connaître, c'est simuler ». De même que « modèle », le mot « simulation » a un sens fort différent suivant qu'on l'emploie dans un contexte scientifique ou qu'on le prenne dans son acception ordinaire. Dans le premier cas, la simulation est une forme particulière de modélisation qui consiste à reproduire, le plus souvent sur ordinateur, le fonctionnement d'un système ; il s'agit donc d'une activité de connaissance. Dans le second, « simulation » rime avec « dissimulation » : on est dans le domaine de la feinte, du faire-semblant, du simulacre. L'aboutissement du *verum factum* dans les sciences de la cognition, c'est d'avoir confondu ces deux sens.

En 1950, Turing, dans un article non moins célèbre que celui de 1936, « Computing Machinery and Intelligence[18] », propose une méthode opératoire pour trancher la question qui agite les esprits : « Les machines peuvent-elles penser ? » Cette méthode a la forme d'un jeu. Turing lui donne significativement le nom de « jeu de l'*imitation* ».

Il aurait pu penser au montage suivant. Le jeu se joue à trois, une machine, un être

humain et un interrogateur qui, ne voyant ni n'entendant les deux premiers, doit déterminer qui est qui en conversant avec eux au moyen, par exemple, d'un téléscripteur. La machine a pour stratégie d'induire l'interrogateur en erreur en se faisant passer pour l'être humain, lequel s'efforce au contraire d'affirmer son identité. La machine aura suffisamment bien *simulé* le comportement de l'être humain si l'interrogateur se révèle incapable de décider. On aura du même coup isolé un niveau de représentation et de communication, autonome par rapport à la matérialité des systèmes cognitifs, où se situent l'intelligence, la pensée, la connaissance, etc.

Le « jeu de l'imitation » est en réalité plus subtil, il comporte une étape préliminaire. Dans celle-ci, l'interrogateur a affaire à un homme et à une femme, le premier tentant de se faire passer pour la seconde. Vient alors seulement la question : que se passera-t-il si l'on remplace l'homme par une machine ? L'introduction d'une étape supplémentaire change sensiblement la nature du « test » : pour dérouter l'interrogateur, la machine doit désormais simuler non simplement le comportement d'un être humain (en l'occurrence une femme[19]), mais bien la capacité de simulation de celui-ci (en la personne de l'homme). La simulation de la faculté de simulation, telle est bien la définition fonctionnaliste de l'esprit.

Une machine arrivera-t-elle un jour à donner le change, protégée ainsi par ce qui s'apparente à un « voile d'ignorance »[20] ? La controverse a moins porté sur ce point que sur l'adéquation même du test. Une machine pourrait fort bien le réussir sans qu'on puisse pour autant en conclure qu'elle pense, ont protesté plus d'un. Parmi eux, John Searle et sa célèbre expérience de pensée, « la pièce chinoise[21] ». Imaginons, suggère-t-il, que moi qui ne comprends rien au chinois, sois enfermé dans une pièce avec, à ma disposition, un réservoir de caractères chinois et un programme, écrit dans ma langue maternelle (l'anglais ou le français), de manipulation desdits caractères que je suis capable d'identifier par leur seule forme. De l'extérieur, on me pose des questions sous forme de séquences de caractères ; grâce aux instructions de mon programme, je leur fournis des réponses sous la forme d'autres séquences, que je fais passer à l'extérieur. Le système entrée-sortie fonctionnant de cette manière est, placé sous le voile du jeu de l'imitation, indistinguable d'un Chinois qui répondrait aux questions qu'on lui pose en *comprenant* ce qu'il fait. Il est cependant clair que pas plus moi que le système constitué par la pièce ne pouvons prétendre avoir pensé ce que nous faisions. La manipulation réglée de symboles pourra tout au plus *simuler* la production et la communication de sens ; elle ne pourra jamais les *réaliser*. Comme Searle l'écrit ailleurs, « un modèle informatique d'un phénomène mental n'est pas plus réel que le modèle informatique de tout autre phénomène naturel[22] ».

Ce que le philosophe américain refuse ici est beaucoup plus que le simple « test de Turing ». C'est toute la progression du *verum factum* dans l'aventure scientifique occidentale, qui a conduit à identifier pensée et production de modèles. Car si l'on admet cela – si l'on accepte que la pensée est (toujours) déjà simulation[23] –, pourquoi refuserait-on à la simulation de la simulation la qualité de pensée ?

John Searle ne voit pas, ou refuse de voir, ce sur quoi nous allons insister dans ce livre. La révolution copernicienne introduite par la « science de l'esprit » aura réalisé une *déconstruction* de la métaphysique de la subjectivité qui va beaucoup plus loin que la démarche philosophique qui s'est donné ce nom. En matière de chinoiseries, on peut préférer, à tout prendre, celles que mitonne son adversaire Jacques Derrida[24] . Selon

Vincent Descombes, l'auteur de *L'Écriture et la différence* observe quelque part que « la seule façon de faire semblant de parler chinois quand on parle à un citoyen chinois, c'est de lui adresser la parole *en chinois*. Par conséquent, dans cet ordre de l'énonciation, la simulation est simulation de simulation (pour faire semblant, je fais vraiment : j'ai donc seulement feint de feindre)[25] ».

2

Une parente mal aimée

« Une science dynamique n'a que faire de son passé, elle va de l'avant. »

Marvin MINSKY[1]

La révolution logique des années trente allait creuser des sillons profonds et variés dans l'histoire de la pensée. Au début des années quarante, l'un d'entre eux fit se regrouper un petit nombre de mathématiciens, d'ingénieurs et de neurobiologistes. Résolus à fonder une science de l'esprit, ils allaient se donner, à partir de 1947, le nom de code de « cybernétique ». Ce livre part de l'hypothèse que ce qu'on appelle aujourd'hui les « sciences cognitives » trouve son origine dans le mouvement cybernétique.

Cette hypothèse ne va pas de soi, et ce pour deux raisons au moins. La première, c'est que la prolifique cybernétique a eu bien d'autres rejetons que les sciences cognitives ; les membres de la fratrie sont si dissemblables qu'on ne saurait leur trouver un air de famille : ils ne se reconnaissent même pas entre eux. On a oublié que la cybernétique, dans ses beaux jours, suscita les plus grands enthousiasmes et les plus folles espérances. Son projet théorique, idéologique et technique a façonné notre époque comme nul autre. Il ne faut donc pas s'étonner que sa lignée soit nombreuse et variée. Elle aura, en vrac et sans souci d'exhaustivité : introduit la conceptualisation et le formalisme logico-mathématiques dans les sciences du cerveau et du système nerveux ; conçu l'organisation des machines à traiter l'information et jeté les fondements de l'intelligence artificielle ; produit la « métascience » des systèmes, laquelle a laissé son empreinte sur l'ensemble des sciences humaines et sociales, de la thérapie familiale à l'anthropologie culturelle ; fortement inspiré des innovations conceptuelles en économie, recherche opérationnelle, théorie de la décision et du choix rationnel, théorie des jeux, sociologie, sciences du politique et bien d'autres disciplines ; fourni à point nommé à plusieurs « révolutions scientifiques » du XXe siècle, très diverses puisqu'elles vont de la biologie moléculaire à la relecture de Freud par Lacan, les métaphores dont elles avaient besoin pour marquer leur rupture par rapport à des paradigmes établis.

Les sciences cognitives, par souci de pureté et esprit de conquête, rêvent de faire table rase du fouillis conceptuel qui règne dans les sciences de l'homme : elles ne sauraient reconnaître leur place au milieu de ce fatras.

Il y a un second obstacle à l'attribution de parenté que je propose : la résistance, pour dire le moins, du descendant à l'admettre. Les sciences cognitives ont honte de la cybernétique, lorsqu'elles reconnaissent avoir quelque chose en commun avec elle. J'aurai plusieurs fois l'occasion d'illustrer cet ostracisme, que l'on retrouve dans la plupart des domaines que la cybernétique aura marqués de son influence. Je n'insiste donc pas ici, sinon pour dire que c'est ce rejet qui m'a incité à entreprendre l'enquête présente.

Dans le monde des hommes, ceux qui rougissent de leurs parents ont peu de chances de se développer harmonieusement. Il n'y a aucune raison qu'il en aille différemment dans le monde des idées. La volonté d'oubli de sa propre histoire, propre à l'optimisme scientiste, est le moyen le plus sûr de se condamner à la répéter en refaisant les mêmes erreurs. Or la cybernétique en a commis, et de graves, les premiers à l'affirmer sont précisément ceux de ses enfants qui voudraient bien se cacher qu'ils sont issus de sa matrice.

Une « scienza nuova » ?

L'histoire commence en 1943. Fixer une origine est toujours arbitraire, mais ici moins qu'ailleurs. En cette année faste pour la science de l'esprit paraissent indépendamment deux articles dont les auteurs, trois pour l'un, deux pour l'autre, constitueront le noyau dur du mouvement cybernétique. Ces deux textes ouvrent sur des programmes de recherche qui, réunis, définiront le projet de la nouvelle discipline.

Le premier est signé Arturo Rosenblueth, Norbert Wiener et Julian Bigelow, et s'intitule « Behavior, Purpose and Teleology ». C'est la revue *Philosophy of Science* qui le publie, dans le numéro 1 de son volume 10, en ce mois de janvier 1943. Wiener, qu'il faut définir avant tout comme un mathématicien appliqué – un « fabricateur de modèles » –, travaillait pendant la guerre avec Bigelow, un jeune ingénieur, aux problèmes théoriques posés par la défense anti-aérienne. Cette recherche, réalisée au MIT, était dirigée par Warren Weaver, lequel allait cosigner avec Claude Shannon leur fameuse *Mathematical Theory of Communication* (1949)[2] . Bigelow, plus tard, sera recommandé par Wiener à John von Neumann et deviendra, à Princeton, l'ingénieur en chef de celui-ci pour la construction de l'ordinateur maison, le JONIAC, qui jouera un rôle essentiel dans le développement de la bombe H[3].

Le problème central de la défense anti-aérienne, c'est que, la cible étant mobile, il faut prévoir sa position future à partir d'une information partielle sur sa trajectoire passée. Wiener mit au point une théorie probabiliste de prédiction reposant sur des théorèmes d'ergodicité qui devait révolutionner les techniques de traitement du signal. A côté de ce qui allait s'appeler théorie de l'information, les problèmes de la défense anti-aérienne faisaient intervenir l'autre ingrédient de base de la future cybernétique : le concept de boucle de rétroaction (*feedback*), inhérent à toute régulation d'un système sur la base de l'écart observé entre son action effective (*output*) et le résultat projeté (but, *goal*).

Or Wiener, qui était un esprit universel à l'instar de von Neumann, s'intéressait depuis longtemps à la physiologie. Depuis 1933, il participait à un séminaire de la faculté de médecine de Harvard et il y avait fait la connaissance d'un physiologiste mexicain, Arturo Rosenblueth, intéressé comme lui par la philosophie des sciences[4]. Rosenblueth travaillait dans le laboratoire de Walter Cannon et était marqué par la notion d'homéostasie que celui-ci avait dégagée de ses recherches physiologiques en lui donnant un champ d'application très vaste, de l'organisme à la société. Wiener et Bigelow discutèrent avec Rosenblueth de l'analogie qui leur était venue à l'esprit : pouvait-on rapprocher leur conceptualisation de la défense anti-aérienne des processus à l'œuvre dans le mouvement volontaire d'un sujet humain ? Un malade dont le cervelet est lésé et qui essaie de porter à ses lèvres un verre d'eau communique à sa main des mouvements oscillatoires d'amplitude croissante, qui évoquent irrésistiblement le comportement d'une boucle de

rétroaction dérégulée. Les réflexions des trois amis aboutirent à l'article de 1943.

Tout cela est fort connu mais l'examen de l'article réserve néanmoins quelques surprises au lecteur d'aujourd'hui. D'abord par la gaucherie de sa terminologie, qui résulte de ce que les catégories d'information, de communication, d'organisation ne sont pas encore éclaircies. Le *feedback* est décrit comme un retour de l'*énergie* de l'output sur l'input : crime de lèse-cybernétique avant la lettre ! Mais surtout, l'ensemble dégage un parfum de behaviorisme qui ne laisse pas d'étonner, si l'on sait que, de toutes les écoles de psychologie, ce fut celle avec laquelle la cybernétique daigna le moins dialoguer. L'article se donne comme premier objectif de « définir l'étude comportementale (*behavioristic*) des phénomènes naturels et de classer les comportements ». Le comportement d'un objet est défini comme « toute modification décelable *du dehors* » (je souligne). La méthode préconisée ne se préoccupe que des relations de l'objet avec son environnement et, contrairement à son opposée, l'« analyse fonctionnelle », « néglige, dans l'objet, sa structure spécifique et son organisation propre ».

La cybernétique, telle qu'elle se trouve anticipée dans l'article de 1943, traite indéniablement ses objets d'étude comme des dispositifs transformant des messages d'entrée (*input*) en messages de sortie (*output*). C'est une définition que l'on trouvera encore en toutes lettres dans l'ouvrage tardif de Wiener, *God and Golem, Inc.* (1964). Ce qui l'empêche cependant de se réduire à un simple behaviorisme obéissant au schéma stimulus-réponse, c'est précisément la notion de *feedback*. Grâce à ce dispositif, l'objet est capable de changer la relation qu'il établit entre *input* et *output*, entre stimulus et réponse. Pour l'observateur qui choisit de rester à l'extérieur de l'objet, tout se passe comme si celui-ci avait la capacité de modifier sa réponse à un stimulus donné, et ce en vue d'atteindre un certain objectif. Tout se passe donc, apparemment, comme si l'objet était capable de poursuivre une finalité donnée en apprenant à ajuster son comportement au vu des erreurs qu'il commet. Le fondateur du mouvement néocybernétique connu sous le nom de « seconde cybernétique », Heinz von Foerster, exagérera plus tard l'importance de cette rupture avec le behaviorisme, tout en reprochant à la première cybernétique d'être restée très en deçà des possibilités ainsi ouvertes. Il opposera aux « machines triviales » du behaviorisme, asservies à une règle « stimulus-réponse » fixée une fois pour toutes, la très grande richesse de comportements que peut manifester une « machine non triviale », dotée, à l'instar de la machine de Turing, d'un état interne et capable de modifier celui-ci en fonction de l'input et de l'état interne à la période précédente[5]. Une machine non triviale n'est rien d'autre que ce que les logiciens appellent un « automate à états finis ». Pour des valeurs relativement modestes des nombres d'états internes et d'inputs possibles, elle exhibe, par explosion combinatoire, une complication de comportement proprement inextricable. C'est cette complication qui fera dire à von Foerster que la machine non triviale traite de l'*information*, alors que la machine triviale behavioriste réagit à un *signal*.

Il n'empêche. Le lecteur d'aujourd'hui est frappé par le ton behavioriste de l'article de 1943. L'insistance de celui-ci à ne pas prendre en compte l'organisation interne de l'objet a de quoi surprendre tous ceux qui ne connaîtraient de la cybernétique que sa seconde phase, inaugurée par les travaux de von Foerster et de Ross Ashby, culminant dans les théories de l'organisation biologique de Humberto Maturana et de Francisco Varela[6]. L'accent est au contraire alors mis sur la cohérence interne et l'« autonomie » de l'objet, organisme ou machine complexe, et l'on va même jusqu'à réduire ses relations avec

l'environnement à de simples perturbations, en aucun cas porteuses d'information.

Il est intéressant de comprendre ce qui motive cette surprise. L'image que la cybernétique a laissée – aussi bien chez ceux qui se réclament encore d'elle que chez ses détracteurs – est celle d'une *scienza nuova* conquérante, se posant en rivale de la physique et se donnant pour ambition, en substituant la forme à la matière, de mettre fin à la domination séculaire de « la reine des sciences ». Philippe Breton et de nombreux autres commentateurs proposent une interprétation de l'article de 1943 qui s'accorde avec cette image. Ce que ce texte fondateur, selon eux, enjoint de négliger, c'est le « contenu », c'est-à-dire la nature *physique* des constituants et de leurs relations, et cela afin d'abstraire la *forme* de celles-ci. Par exemple, on s'emploie à déceler un même dispositif *formel* de feedback dans un animal et dans une machine, alors même que le premier est matérialisé par des protéines et le second par des tubes électroniques. De là les conclusions de l'article, « qu'une analyse comportementale uniforme est applicable à la fois aux machines et aux organismes vivants ». De là aussi que la catégorie de « téléologie », comprise comme « finalité (*purpose*) réglée par rétroaction négative », peut s'appliquer aussi bien aux premières qu'aux seconds. Ainsi entendu, selon Philippe Breton, ce texte « marque une rupture essentielle avec les conceptions de la science moderne, rupture dont les effets se perpétuent encore largement dans la pensée actuelle[7] ».

C'est beaucoup dire. Les fondateurs de la cybernétique n'avaient pas conscience de bâtir une *scienza nuova*. La révolution que constitua, dans l'histoire de la pensée, l'avènement de la science nouvelle était loin derrière eux, comme nous l'avons vu au chapitre précédent. Le fait d'abstraire la forme des phénomènes et, par là même, de se rendre capable de repérer des isomorphismes entre domaines différents, c'est la démarche modélisatrice par excellence, c'est la démarche scientifique même. Si elle permet de proposer une théorie unifiée de la machine et du vivant quant à la catégorie de finalité, pensée en termes mécanistes et rebaptisée « téléologie », c'est là certes une avancée spectaculaire de la science quant à son extension ; ce n'est en aucune façon une rupture. Les cybernéticiens, en particulier, n'avaient pas la prétention de rompre avec la physique, ni de la dépasser. C'est au contraire dans son cadre, celui dans lequel ils avaient reçu leur formation scientifique, qu'ils entendaient situer les notions nouvelles. Nous aurons l'occasion de le vérifier à de multiples reprises.

La volonté affichée de rester à l'extérieur de l'objet et d'ignorer son « contenu » cache en vérité autre chose. La lecture des Actes des conférences Macy, et ce qu'elle nous apprend au sujet des attitudes profondes des auteurs de l'article de 1943, dicte une réponse qui s'accorde d'ailleurs mieux avec la lettre de cet article. On pourrait la lire dans les nombreuses passes d'armes qui opposèrent les cybernéticiens à la psychanalyse, à propos de la notion d'inconscient, mais c'est en toutes lettres qu'on la trouve exprimée à l'occasion d'un heurt entre Rosenblueth, appuyé par Bigelow, et un spécialiste de la communication animale, Herbert Birch. Cela se passe lors de la huitième conférence Macy, en 1951. Birch vient de passer en revue divers cas de ce qu'on appelle la « communication » chez les insectes et les vertébrés inférieurs. Le message qu'il a martelé au long de ces exemples est que derrière des comportements impressionnants de complexité apparente, et encore mal compris, il *ne* faut voir *que* des comportements, précisément, stéréotypés, et qu'il serait vain d'en appeler à des capacités d'intelligence et de compréhension propres à une « vraie » communication. Pour mieux convaincre, il a pris l'exemple extrême suivant, qui ne lui sera pas pardonné. La coquille Saint-Jacques

est un des mets favoris de l'étoile de mer. Toute étoile de mer située à proximité d'une coquille Saint-Jacques provoque chez celle-ci une réaction de fuite. Dira-t-on que la première a « communiqué » avec la seconde ? Si oui, continuera-t-on à le dire lorsqu'on remplace l'étoile de mer par un extrait de bouillie d'étoile de mer, qui produit exactement le même effet ? Birch termine son exposé par l'examen des caractéristiques distinctives des « vraies » communications que l'on n'observe, selon lui, qu'à partir des mammifères supérieurs et qui culminent chez l'homme : elles impliquent anticipation, intentionnalité, symbolisation, et requièrent des capacités d'apprendre, de percevoir, de tisser des relations sociales. Rosenblueth explose. Ces distinctions sont privées de sens, explique-t-il en substance. Elles reposent sur des notions : intelligence, conscience, mémoire, apprentissage, anticipation, intentionnalité, contenu, qui renvoient à ce qui se passe dans l'esprit (*mind*) des sujets et ne peuvent faire l'objet d'aucune mesure : elles n'ont aucune pertinence pour le problème de la communication. Les seuls contacts que nous puissions avoir avec les autres passent par leur *comportement* : c'est la seule chose que nous puissions voir, que nous puissions juger, et qui peut avoir un effet sur nous. Certes, concède-t-il, lorsque nous décrivons le comportement des êtres vivants, qu'il s'agisse d'un humain ou d'un organisme inférieur, voire d'une machine, il nous arrive, c'est même parfois indispensable, de nous servir d'une terminologie « mentaliste » comme si nous postulions en eux un « esprit ». Cela n'est pas gênant à une double condition : que nous y voyions une simple commodité de langage, et que nous comprenions que cela n'a pas de pertinence pour la question qui nous intéresse : la communication. En ce sens, nous pouvons très bien dire d'une machine qu'elle a une « mémoire », qu'elle « apprend ». Cela ne tire pas à conséquence, tant que nous nous référons seulement à quelque chose d'objectif, de mesurable, qui n'a aucun rapport avec notre expérience intérieure[8].

La rupture avec le behaviorisme doit donc être très fortement relativisée. La science de l'esprit que la cybernétique entend bâtir est, dans la terminologie d'aujourd'hui, résolument « éliminationniste ». Les « états mentaux » que la psychologie ordinaire ou « naïve » invoque pour rendre compte des comportements, comme les croyances, les désirs, la volonté, les intentions, etc., doivent être bannis de l'explication scientifique. La cybernétique de Wiener et de Rosenblueth est foncièrement non mentaliste. Cela, on ne l'a pas toujours assez compris, à commencer par les cybernéticiens français qui ont eu connaissance de l'article de 1943 dans la traduction qu'en ont donnée les *Études philosophiques* dans leur numéro 2 de 1961. Le titre « Behavior, Purpose and Teleology » y devenait : « Comportement, intention, téléologie ». La traduction de *purpose* par intention est un contresens parce qu'elle relève d'un mentalisme qu'il s'agit précisément de chasser, *purpose* désignant ici une finalité *non intentionnelle*. Heinz von Foerster lui-même reprochera souvent à la première cybernétique, et à l'intelligence artificielle qui en prit la suite, d'avoir parlé des machines en termes anthropomorphiques. C'est être aveugle au fait que ce n'est pas la machine que les premiers cybernéticiens dotent d'humanité – c'est l'humain que, délibérément, ils assimilent à la machine. La distinction, si fondamentale dans l'histoire de la physiologie, entre comportement volontaire et comportement réflexe perd avec eux tout son sens, de même que la différence entre conscience et inconscient. L'assimilation de l'homme à la machine n'est cependant pas conçue comme une réduction. Car la machine est le *modèle* – dans les deux sens du mot.

Mécaniser l'humain

Le second article fondateur publié en 1943 est signé par le neuropsychiatre Warren McCulloch et le mathématicien Walter Pitts. Son ambition philosophique est considérable puisqu'il ne s'agit de rien moins que de donner une base purement neuro-anatomique et neurophysiologique au jugement synthétique *a priori*, et de fonder ainsi une neurologie de l'esprit.

L'article de McCulloch et Pitts radicalise la démarche de Wiener et de ses collaborateurs sur deux plans. Alors que ceux-ci dénient toute réalité à l'esprit et réduisent l'évocation de ses facultés à de simples commodités de langage, McCulloch part en quête des mécanismes matériels et logiques qui incarnent l'esprit (*Emhodiments of Mind* est le titre qu'il choisira pour le recueil de ses principaux articles, en 1965), et n'hésite pas à se demander « pourquoi l'esprit est dans la tête[9] ». Il pourrait donc sembler qu'il y ait une opposition très forte entre le « mentalisme » de McCulloch et la critique behavioriste du mentalisme développée par Wiener. Lors de ses nombreuses confrontations avec les gestaltistes, il arriva même à McCulloch d'être accusé de « dualisme ». Wolfgang Köhler tenta d'argumenter que la démarche de celui-ci le conduisait, malgré lui, à attribuer une autonomie à un monde des « apparitions », des faits mentaux, à un monde psychologique, avec ses déterminations propres[10]. McCulloch, en vérité, introduit l'« étude comportementale des phénomènes naturels », chère à Wiener, Rosenblueth et Bigelow, à l'intérieur du cerveau. Certes, le « contenu » de cela même qui est capable de comportement est maintenant considéré comme justiciable d'une démarche scientifique, mais ce contenu se décrit lui-même en termes de comportements d'unités plus petites, à l'« intérieur » desquelles il n'est pas question de pénétrer et que l'on n'envisage que dans leurs relations avec leur environnement, c'est-à-dire comme opérateurs transformant des inputs en Outputs : les neurones. Ce n'est donc pas d'opposition qu'il faut parler, mais bien de radicalisation puisque l'approche comportementale et communicationnelle est transposée à un niveau logique inférieur.

L'autre élément de radicalisation concerne les rapports entre organismes et machines. La cybernétique telle que l'a popularisée Wiener se présente comme la science des *analogies* maîtrisées entre organismes et machines[11] . On peut résumer la position de McCulloch par cette citation, extraite d'un texte de 1955 : « Plus nous apprenons de choses au sujet des organismes, plus nous sommes amenés à conclure qu'ils ne sont pas simplement analogues aux machines, mais qu'ils *sont* machines[12]. » La différence est double. Elle est d'abord de point de vue. Wiener raisonne en mathématicien appliqué, il lui suffit d'établir un isomorphisme mathématique pour conclure à l'analogie. C'est une prise de position ontologique qu'en regard défend McCulloch. Par ailleurs – mais les deux choses évidemment sont liées – les machines dont parle Wiener sont de « vraies » machines, des artefacts, des machines artificielles, des objets techniques. Pour McCulloch, la machine est un être logico-mathématique incarné dans la matière de l'organisme ; c'est, si l'on veut, une « machine naturelle », ou bien une « machine logique », nature et logique étant ici parfaitement équivalentes l'une à l'autre. C'est donc sans incohérence aucune que McCulloch prendra fréquemment ses distances par rapport aux analogies parfois légères établies par ses collègues entre automates artificiels et automates naturels – ces analogies qui feront tant pour discréditer la cybernétique aux yeux des scientifiques « sérieux », c'est-à-dire disciplinaires.

Une confrontation célèbre, dont nous aurons l'occasion de reparler, eut lieu en 1948 entre McCulloch et John von Neumann, lors d'un colloque de l'institut californien de

technologie (CalTech) connu sous le nom de « symposium Hixon ». John von Neumann y présenta sa fameuse conférence « The General and Logical Theory of Automata », en prenant grand soin de marquer les différences aussi bien que les traits communs entre automates naturels et artificiels. L'attitude de McCulloch est révélatrice : il se retrancha significativement dans le camp des biologistes et des psychologues, comme pour mieux reléguer von Neumann dans celui des mathématiciens appliqués et des ingénieurs. L'enjeu de cette prise de distance est proprement philosophique et porte sur le statut du modèle. Le modèle, pour McCulloch, n'est pas un simple instrument de calcul, dont la valeur serait purement pragmatique, du type « est-ce que ça marche ? ». Il a une réalité ontologique[13]. Le texte de 1955 que nous avons cité plus haut se conclut d'ailleurs ainsi : « Les machines faites de main d'homme ne sont pas des cerveaux mais les cerveaux sont une variété, très mal comprise, de machines computationnelles. La cybernétique a contribué à effondrer la muraille qui séparait le monde magnifique de la physique du ghetto de l'esprit. » Notons que cette citation, à elle seule, suffit à faire justice de deux idées fausses largement répandues au sujet de la cybernétique, et que nous avons déjà eu l'occasion de relever. Il est clair qu'il ne s'agit pas ici d'humaniser la machine, mais bien de mécaniser l'humain ; il est non moins clair que la physique, loin d'être la rivale à abattre, est le modèle à imiter.

McCulloch a souvent rappelé l'itinéraire singulier qui l'a conduit, associé au mathématicien et logicien de génie qu'était Walter Pitts, à concevoir l'article de 1943. Parti d'une interrogation philosophique qui, toute sa vie, le tenaillera : comment savons-nous ce que nous savons, et comment désirons-nous ce que nous désirons ? – interrogation qu'il précisera plus tard sous la forme de la question : « What is a Number, that a Man may know it, and a Man, that he may know a Number[14] ? » –, il traversa les disciplines les plus variées, toujours insatisfait des réponses qu'elles donnaient à sa quête, de la philosophie à la psychologie, de la psychologie à la neuropsychiatrie et à la neurophysiologie, et enfin de celles-ci à la logique. Dès les années vingt, encore psychologue, sous l'influence de Whitehead et de Russell, mais peut-être aussi de Démocrite, il cherche à construire un calcul propositionnel dont les propositions atomiques sont des événements psychiques, les psychons, localisés dans le temps et ayant les propriétés suivantes : chaque psychon a la propriété de tout ou rien : ou il se produit, ou il ne se produit pas ; chaque psychon implique son antécédent temporel et intervient dans la détermination des psychons subséquents ; il est possible de combiner les psychons de façon à produire des propositions plus complexes concernant leurs antécédents. En 1929, il lui vient à l'esprit que ces atomes psychiques sont peut-être les impulsions électriques que chaque neurone envoie (ou non) à ses voisins.

L'idée que le calcul logique des propositions, l'« algèbre de Boole », est matérialisable par des circuits et des relais électriques nous est devenue familière, voire banale, à nous qui vivons dans un monde d'ordinateurs. Nous avons du mal à imaginer aujourd'hui le choc intellectuel que fut la découverte de cette idée pour ceux qui l'éprouvèrent. Herbert Simon a avoué que le cours de sa carrière intellectuelle se décida à ce moment-là[15] . Nul ne s'étonne, aujourd'hui, de voir le cerveau, à quoi nous devons, pensons-nous, nos facultés logiques, comparé à un calculateur numérique, à un *digital computer*. Il fut par ailleurs un temps, s'en souvient-on ? où l'ordinateur se nommait « cerveau électronique ». Cette double métaphore, la légende veut que ce soit la cybernétique qui l'ait inventée – à la manière, en quelque sorte, des analogies wienériennes entre organismes et machines.

C'est ainsi que Jean-Pierre Changeux, dans son étude historique, peut écrire, comparant – non sans s'entourer de quelques précautions – le cerveau à une « machine cybernétique, à un ordinateur[16] » : « La cybernétique, avec Wiener, reprend cette thèse [la thèse de La Mettrie, selon laquelle "l'homme est une machine" – JPD]. L'homme n'a plus un cerveau comparable à la mécanique d'un automate ou d'une horloge, mais *ressemble à* et *fonctionne comme* un ordinateur[17]. »

Il est donc important de rappeler que les choses ne se sont pas du tout passées ainsi. La première confusion à lever porte sur le nombre des acteurs : ils ne sont pas deux, l'organisme et la machine, mais bien trois : l'organisme dans sa structure (le cerveau), l'organisme dans sa fonction (l'esprit) et la machine, cette dernière se dédoublant elle-même en machine logique (machine de Turing ou machine de McCulloch et Pitts) et machine artificielle, matérielle (l'ordinateur). Ce troisième acteur, dans sa première incarnation, joue le rôle du modèle, et tel le crâne de Yorick, c'est lui qui est le centre de l'action. Celle-ci comporte trois moments.

Cerveau/esprit/machine

Le premier, c'est l'assimilation de l'esprit à une machine. Machine logique, d'abord. Au chapitre précédent, nous avons montré comment l'aventure scientifique occidentale avait abouti à ce point d'orgue, au terme d'un long *crescendo* culminant dans la révolution logique des années trente, ponctuée par les théorèmes de Gödel et de Turing, et la thèse de Church-Turing. Il faut ajouter une voix à cette déjà riche partition, celle de Claude Shannon, l'un des futurs participants aux conférences Macy. Dans un article paru en 1938 dans une revue technique d'électricité, celui qui sera plus tard l'un des fondateurs de la théorie de l'information propose une théorie logique des circuits électriques, et plus spécialement des réseaux d'impédance à relais ou à commutateurs. La nouveauté de cette étude est double. Ces réseaux ont bien évidemment déjà fait l'objet de modélisations mathématiques, mais selon une mathématique des quantités : l'innovation réside dans le recours à l'outil logique, à savoir le calcul des propositons. Par ailleurs, et inversement, il existe avant l'étude de Shannon une longue tradition de recherches visant à résoudre des problèmes de logique au moyen de dispositifs matériels mécaniques. On peut la faire remonter au début du XIXe siècle, et le « piano logique » de Jevons, dans les années 1870, en constitue l'un des moments forts[18] . La discontinuité introduite par Shannon réside dans la finalité même du rapprochement qu'il effectue entre machine et logique[19]. La démarche perd le caractère pragmatique de la résolution de problèmes, elle devient essentiellement théorique. Cette théorie a beau être celle d'un ingénieur se fixant des buts techniques – il faut ici concevoir la synthèse de réseaux capables de réaliser un comportement donné –, elle ouvre inévitablement des perspectives philosophiques. Car si la logique se révèle capable de fournir un *modèle* de fonctionnement d'un réseau électrique, c'est donc, par le basculement que le recours au modèle provoque inévitablement, que la logique se réalise, s'incarne dans un mécanisme matériel. Machine artificielle ou matérielle, d'un côté, logique comme machine, de l'autre, sont liées par un rapport d'identité. Comme, par ailleurs, on avait déjà à l'époque tous les éléments pour se convaincre que l'esprit est une machine logique, Shannon aurait pu conclure, par transitivité, que l'esprit est une machine matérialisable – et donc, peut-être, réplicable par

l'artifice de l'homme. L'ingénieur, en lui, répugnait sans doute à de telles enjambées. D'autres n'allaient pas tarder à franchir le pas.

L'impact qu'a eu le travail de Shannon sur la formation du modèle de McCulloch et Pitts est mal connu. Au cours des conférences Macy, il n'en est jamais question, le Shannon qui intervient dans les discussions étant essentiellement celui qui, entre-temps, a mis au point la théorie de l'information. Il est évident, en revanche, que la thèse Turing exerça sur McCulloch et Pitts, ainsi que sur quelques autres participants aux conférences Macy, une influence considérable. L'article de 1943 n'aurait pas vu le jour sans l'*impetus* que les deux collaborateurs en reçurent. Les Actes des conférences Macy, ainsi que bien d'autres textes écrits par les cybernéticiens à l'époque, révèlent d'ailleurs que, malgré leur formation scientifique et logique, ils commettent fréquemment l'erreur, propagée par les idéologues, de présenter la thèse de Turing comme un résultat susceptible d'être démontré et, de fait, démontré. Lors de la neuvième conférence Macy, en 1952, Pitts expose ce que doit être selon lui la responsabilité pédagogique du cybernéticien. En construisant des machines capables de remplir telle ou telle fonction qui pouvait sembler être l'apanage de l'esprit ou d'une mystérieuse substance psychique non matérielle, il joue un rôle éducatif inappréciable. Mais, précise Pitts, se référant à la thèse ou au théorème de Turing, on ne sait trop, *pour nous* l'affaire est depuis longtemps entendue, et cela une fois pour toutes ; aussi bien ce travail de persuasion au coup par coup, fonction psychique après fonction psychique, ne vaut-il que pour le « reste du monde »[20]. Cette profession de foi est d'autant plus remarquable que c'est l'un des très rares moments, dans les conférences Macy, où le « nous » est employé pour désigner la communauté cybernéticienne. A ressentir plus de quarante ans après par la lecture le frémissement de Pitts, on perçoit intensément que, pour lui, la thèse de Turing était comme un trésor secret dont il s'attribuait la garde, une manne trop précieuse pour être dissipée à la légère.

Le deuxième moment est précisément constitué par l'article de 1943 de McCulloch et Pitts, intitulé « A Logical Calculus of the Ideas Immanent in Nervous Activity ». Il démontre en principe l'existence d'une machine logique équivalente à celle de Turing (au sens que tout ce que peut faire l'une, l'autre peut le faire et réciproquement), qui peut être considérée dans sa structure et dans son comportement comme une idéalisation de l'anatomie et de la physiologie du cerveau. Ce résultat, dans l'esprit de McCulloch tout au moins, constitue une avancée décisive, puisque ce n'est plus seulement le cerveau dans sa fonction (l'esprit), mais aussi le cerveau dans sa structure, le cerveau matériel, naturel, biologique, qui est assimilable à un mécanisme, et plus précisément à une machine de Turing. C'est ainsi que le cybernéticien pense résoudre le vieux problème de l'âme et du corps, ou, dans ses termes, de l'« *embodiment of mind* ». Le cerveau et l'esprit sont Fun et l'autre une machine, et c'est la même machine. Le cerveau et l'esprit, donc, ne font qu'un.

Les connaissances neuro-anatomiques et neurophysiologiques de l'époque n'ont pas la précision de celles d'aujourd'hui, c'est évident, mais surtout elles sont moins assurées. Rappelons que l'introduction de la microscopie électronique date des années cinquante. Certes, depuis le début du siècle, l'école espagnole, avec Ramon y Cajal, puis avec Lorente de No' (lequel jouera un rôle important dans le dévelopement de la cybernétique), a puissamment contribué à accréditer la thèse que le fonctionnement du système nerveux central est assimilable à celui d'un réseau de communications entre neurones. Mais les mécanismes de transmission synaptique sont encore mal débrouillés, puisque c'est seulement en 1952 que Hodgkin et Huxley vont élucider complètement les mécanismes

ioniques du déclenchement et de la propagation de l'influx nerveux. Surtout, les thèses « continuistes », qui présentent le cerveau comme un milieu continu dont le fonctionnement ne peut être appréhendé que comme un tout indivis, sont encore vivaces, et les cybernéticiens auront à les affronter en la personne du gestaltiste Köhler et aussi, dans une moindre mesure, en celle de Lashley – le grand neurophysiologiste de l'époque, dont nous savons que le bouillant McCulloch avait pour lui un grand respect mêlé de crainte[21] . Il faut donc à McCulloch un certain courage pour présenter, en 1943, un modèle du cerveau sous la forme d'un réseau de neurones idéalisés. Chaque neurone reçoit ou non de ses voisins des impulsions, et lui-même est mis à feu si et seulement si une somme pondérée[22] des 1 ou 0 qui codent l'existence ou l'absence d'une impulsion dans les synapses afférentes est supérieure à un certain seuil, dit seuil d'excitation[23]. En d'autres termes, ceux du calcul des propositions, chaque neurone idéalisé est un calculateur arithmétique élémentaire, qui compute une fonction booléenne à seuil. Le cerveau tout entier est représenté comme un réseau d'interconnexions entre de tels calculateurs.

De 1943 à 1963, seul ou en collaboration avec Pitts, McCulloch allait continuer à explorer les possibilités de sa machine logique, convaincu qu'il était, par la thèse de Turing, de sa capacité à faire preuve de toutes les facultés attribuées à l'esprit : perception, pensée, mémoire, formation des concepts, connaissance et reconnaissance des universaux, volonté et même conscience, si l'on définit celle-ci, à la suite de Spinoza, comme idée des idées. Il devait s'intéresser en particulier aux propriétés résultant de l'existence dans le réseau de circuits fermés, cycles ou boucles. Ces boucles avaient été postulées par le neurologue Lawrence Kubie dès 1930, dans un article paru dans la revue *Brain* sous le titre « A theoretical application to some neurological problems of the properties of excitation waves which move in closed circuits[24] ». Kubie collaborait alors avec Sherrington, dont les travaux sur les mouvements involontaires et les réflexes faisaient autorité. En 1938, Lorente de No' mit en évidence expérimentalement l'existence de cycles dans le système nerveux. En 1941, Kubie, devenu psychanalyste, conjectura que le soubassement neurophysiologique des névroses réside dans ces circuits fermés réverbérants, où les séquences d'impulsion viennent se laisser piéger en une ronde sans fin[25]. Jacques Lacan, dont nous reparlerons, eut connaissance de cette lecture cybernétique de la compulsion, ou automatisme, de répétition, le fameux *Wiederholungszwang* freudien. Son intérêt dans les années cinquante pour la cybernétique trouve vraisemblablement là son origine.

De son propre aveu, McCulloch fut très influencé par l'article de 1930 de Lawrence Kubie. La chose mérite d'autant plus d'être notée que les Actes des conférences Macy mettent en scène de façon très vivante les nombreuses passes d'arme entre McCulloch et le Kubie psychanalyste qui, comme seul représentant à ces conférences de la discipline honnie des cybernéticiens, était plus souvent qu'à son tour leur tête de Turc[26] . Les cycles de neurones jouent un rôle important dans la conceptualisation de McCulloch, et le thème de la « causalité circulaire » est évidemment l'un de ceux qui le rapprocheront de Wiener et de ses collaborateurs. Ces cycles, pense McCulloch en 1943 – à condition de ne pas les considérer selon le modèle de l'arc réflexe, qui est un *feedback* négatif et régulateur, mais comme des « circuits réverbérants » à *feedback* positif –, sont cela même qui capte et place hors du temps les séquences d'événements qui atteignent le système nerveux à travers les organes des sens. C'est donc sur eux qu'il faut bâtir une théorie de la mémoire

et de la formation des idées générales. Commentant l'article de 1943 sur ce point, Jean Mosconi peut écrire : « Le mécanisme de conservation d'une trace du passé par réverbération dans les cycles de neurones est bien, du même coup, arrachement de l'événement à sa localisation temporelle précise pour le placer dans un présent indéfiniment renouvelé. Dans les vieilles querelles psychologiques sur les relations entre l'abstraction et l'oubli s'introduit ainsi quelque lumière, par référence non seulement à un substrat physiologique mais aussi à un contenu logique rigoureux et positif : puisque le phénomène considéré correspond au passage de simples formules à constantes (tel événement s'est produit l'instant to) à des formules quantifiées (par exemple : il y a un instant passé où tel événement s'est produit), donc à un niveau supérieur de complexité logique et de capacités expressives[27]. »

Au cours des conférences Macy, McCulloch sera amené à reconnaître l'insuffisance du modèle de 1943 par rapport à l'une de ses principales ambitions : rendre compte de la capacité de l'esprit à former des universaux. Cela ne le détournera pas de sa machine logique, loin de là. Il cherchera à la complexifier par tous les moyens accessibles à son imagination. Il tentera d'abord de la doter de capacités d'apprentissage, telles que trace soit gardée, dans le schéma des connexions et dans les seuils d'excitation, des événéments passés. Toujours convaincu de la nécessité de rester proche de la réalité biologique, il introduira, avec Pitts, de l'aléatoire dans ses réseaux, de façon à les rendre capables de fonctionner en présence d'erreurs et de « bruit » affectant les calculateurs élémentaires et leurs connexions. Dès octobre 1946, lors de la deuxième conférence Macy, Pitts présentait le résultat de ses premières réflexions sur les réseaux aléatoires. Cela suffit pour lancer von Neumann sur la voie qui devait le conduire à construire sa logique probabiliste[28] . La même année, McCulloch comparait l'apprentissage dans des réseaux de neurones aléatoires aux mécanismes de l'aimantation[29].

Apprentissage, rôle du hasard, rapport aux modèles de la physique des systèmes désordonnés : que McCulloch ait eu l'intuition que son modèle pouvait et devait être développé selon ces voies doit en toute justice être porté à son actif, même si Pitts et lui-même en restèrent aux premiers balbutiements. Cela pose la question, fort controversée, du sens et de la valeur de ce travail de pionnier, et de ce qu'il en reste aujourd'hui.

Ceux qui, *post mortem*, veulent du bien à Warren McCulloch préfèrent ne retenir de son apport que sa contribution à la logique formelle et à la théorie de la calculabilité. Parmi eux, on trouve deux de ses anciens collaborateurs, Marvin Minsky et Seymour Papert, plus connus comme pionniers de l'intelligence artificielle, cette discipline qui rallumera, pour le meilleur ou pour le pire, le flambeau éteint de la cybernétique. A considérer leur empressement filial à détourner de leur ancien mentor les accusations fréquemment portées contre la cybernétique, on songe à la plaisanterie bien connue : « Mon Dieu, protégez-moi de mes amis, je me charge de mes ennemis » ; ou, plus sérieusement, au mot de Sartre : « Être mort, c'est être en proie aux vivants. »

La logique joue certes un rôle essentiel dans le modèle de neurones idéalisés de McCulloch et Pitts. Le rapprochement avec la démarche de Shannon est ici pertinent[30] . Avant McCulloch, on avait déjà tenté d'étendre à la biologie, et, dans le cadre de celle-ci, à la théorie du système nerveux, la méthode qui avait si bien réussi en physique : la modélisation mathématique. Cela ne s'était pas fait sans mal, la plupart des neurophysiologistes résistant fortement au projet d'une neurologie mathématisée. L'un d'entre eux, cependant, N. Rashevsky, entreprit dès les années trente de poser les

prolégomènes à une « biophysique mathématique ». Parmi ses élèves, à la fin de la décennie, on comptait le jeune prodige Walter Pitts[31] . L'influence de la physique était si forte que les premiers modèles conservaient les outils et l'esprit même de la modélisation physico-mathématique. Pour rendre compte d'un phénomène macroscopique – ici, les propriétés du système nerveux –, on intègre un système d'équations différentielles qui décrivent les interactions quantitatives entre éléments constituants – ici, les neurones. Cela suppose, évidemment, que l'on puisse traiter les quantités comme des grandeurs différentiables, donc continues. La nouveauté introduite par McCulloch et Pitts fut, ainsi que Shannon l'avait fait dans le domaine des circuits électriques, de prendre au sérieux le caractère de « tout ou rien » de la mise à feu neuronale, et de proposer en conséquence un modèle de type logique. Rashevsky perçut le caractère révolutionnaire de l'article et décida de le publier dans sa revue, le *Bulletin of Mathematical Biophysics*[32].

Peut-on dire pour autant que McCulloch et Pitts firent essentiellement un travail de logicien ? Si tel était le cas, ils ne mériteraient certainement pas la palme. Leurs démonstrations se sont révélées fausses sur des points essentiels, et elles durent être corrigées par Kleene en 1956[33], et complétées par Arbib en 1961. Mais surtout, le réseau de neurones à la McCulloch et Pitts n'est pas en soi équivalent à la machine de Turing, et la classe des fonctions arithmétiques que de tels réseaux permettent de calculer est strictement incluse dans la classe des fonctions récursives, dont on a vu qu'elle est équivalente à la classe des fonctions calculables par machine de Turing. Comme le fait remarquer, non sans arrière-pensées, Seymour Papert dans son « Introduction » de 1965 à *Embodiments of Mind*., ce qu'ont réalisé McCulloch et Pitts d'un point de vue strictement logico-mathématique, c'est d'introduire un nouveau concept de calculabilité – que Kleene associera au concept d'automate *fini*. Comme McCulloch et Pitts le reconnaissent eux-mêmes, pour que leurs réseaux aient les mêmes capacités computationnelles que les machines de Turing, il faudrait leur adjoindre ces deux éléments essentiels de toute machine de Turing que sont une tête mobile capable de lire, d'écrire, d'effacer des symboles et surtout un ruban (une mémoire) potentiellement infini. Le cerveau étant un organe matériel, donc fini, il ne peut certainement pas computer tout ce que peuvent les machines de Turing. Les cybernétiens furent les premiers à oublier cette restriction capitale, et tout au long des conférences Macy, on ira répétant sur un ton assuré la formule magique : « Tout ce qui peut se décrire complétement, sans ambiguïté, en un nombre fini de symboles, est computable par un réseau de neurones. » Ils se faisaient ainsi complices de la dérive idéologique que nous avons déjà repérée à propos de la machine de Turing.

Les neurones de McCulloch

McCulloch fut le premier à rire du malentendu consistant à le prendre pour un logicien. A la fin de sa vie, il ironisait sur l'abstraction croissante du domaine de recherches que son texte de 1943 avait ouvert. Dès le milieu des années soixante, il regrettait amèrement que ce qu'il avait conçu comme théorie des automates (*theory of automata*), c'est-à-dire théorie de ces objets *réels*, qu'ils soient naturels ou artificiels, que l'on nomme automates et qui *incarnent* les facultés de l'esprit, soit devenue peu à peu une pure théorie mathématique désincarnée : l'automatique (*automata theory*)[34]. Il n'est peut-être pas de texte qui exprime mieux la véritable nature de son projet et le sentiment qu'il avait de sa

mission que celui-ci :

« Clerk Maxwell lui-même, alors qu'il n'avait pas de désir plus grand que celui de connaître le lien qui unit les pensées aux mouvements moléculaires du cerveau, coupa court à sa recherche d'un mémorable : "Le chemin qui y conduit ne passe-t-il pas par le cabinet du métaphysicien, cet antre jonché des ossements des explorateurs précédents, qui est l'objet de la détestation de tout homme de science ?" Répondons paisiblement à la première partie de la question par un "si", à la seconde par un "non", et mettons-nous en route sereinement. Notre aventure est en vérité une grande hérésie. Nous sommes sur le point de concevoir le sujet connaissant [*knower*] comme une machine à calculer [*computing machine*][35]. »

Les plus vieilles questions philosophiques sont désormais à la portée de la démarche scientifique. C'est cela, la bonne nouvelle. Réduire le travail de McCulloch à la logique sous prétexte qu'il propose un modèle de l'esprit et du cerveau qui recourt à un formalisme logique, c'est commettre la même erreur de catégorie que celle qui consisterait à croire que le météorologue ou l'écologiste qui résout un système d'équations différentielles fait œuvre mathématique. C'est, une fois de plus, ne pas comprendre le statut du modèle dans la démarche scientifique. McCulloch entendait faire œuvre de psychologue *et* de biologiste, nous l'avons déjà dit. Plus précisément, son ambition était de faire échapper la psychologie à l'« épistémologie spéculative » et d'en faire une « épistémologie expérimentale » grâce à la biologie. En 1964, il ironise sur ceux qui n'ont pas compris que l'objet de sa recherche est « l'incarnation de l'esprit » (*the embodiment of mind*) et précise : « Les neurones que nous avons postulés [dans l'article de 1943], en dépit de leur extrême simplification, sont néanmoins des neurones matériels, aussi sûrement que les atomes du chimiste sont des atomes matériels[36]. »

Si le neurone idéalisé *est* le neurone matériel et qu'en faire la théorie, c'est faire œuvre de biologie et non de logique, c'est que tous les traits du neurone matériel que le modèle formel ne retient pas ne font pas partie de son *essence*. Pour mieux saisir le sens de cette démarche, on peut, par exemple, se demander ce qu'aurait été l'attitude de McCulloch s'il avait pu disposer de nos connaissances sur le système nerveux. S'il avait su ce que nous savons sur les rôles respectifs de la transmission électrique et de la transmission chimique dans les synapses, sur les mécanismes moléculaires et atomiques de la propagation de l'influx nerveux, en eût-il profité pour spécifier, voire amender son modèle, ou bien l'eût-il rejeté ? Ce que cherchait McCulloch, c'était à *abstraire* du système nerveux central un niveau d'organisation et, à ce niveau, certains traits, tels qu'ils permettent le plus économiquement possible de reconstituer les fonctions observées dans le système réel. Ce niveau, il l'avait trouvé : celui des neurones ; ces traits, il les avait dégagés : ceux de ses neurones idéalisés. La vertigineuse descente de nos connaissances à des niveaux d'organisation de plus en plus élémentaires ne l'eût donc pas affecté, et il ne l'eût pas prise pour un progrès décisif par rapport au problème qu'il se posait.

Sa démarche était donc fonctionnaliste – un fonctionnalisme certes fort éloigné de ce que sera le fonctionnalisme computationnel et représentationnel de l'esprit comme machine de Turing, qui constituera la marque distinctive de l'intelligence artificielle et du courant dominant de la science cognitive à venir. Nous en avons parlé au chapitre précédent et nous y reviendrons au suivant. Voilà pourquoi McCulloch insistait sur le fait que les mécanismes matériels qu'il proposait pour rendre compte des facultés de l'esprit étaient *suffisants*, mais en rien *nécessaires*. Ils étaient tout à la fois possibles, comme le

montraient les connaissances anatomiques et physiologiques de l'époque, et capables de manifester ces facultés[37]. Mais d'autres « incarnations », ou d'autres théories de ces incarnations, eussent été concevables. C'est déjà l'idée de « réalisabilité multiple », dont nous avons vu qu'elle est solidaire de la pratique de la modélisation. Or cette idée assure l'émancipation du modèle, lequel peut dès lors voler de ses propres ailes…

Si quelqu'un comprit fort bien cela à l'époque, c'est von Neumann. Parlant de l'article de McCulloch et Pitts de 1943, il écrit : « Ils voulaient traiter des neurones. Ils décidèrent qu'ils ne souhaitaient pas s'attacher aux caractéristiques physiologiques et chimiques d'un neurone réel dans toutes leurs complexités. Ils ont eu recours à ce que l'on nomme en mathématiques la méthode axiomatique, qui consiste à énoncer un certain nombre de postulats simples et à *ne pas se préoccuper des moyens dont la nature se sert pour réaliser l'objet en question*[38]. »

Une bonne illustration de l'interprétation qui précède nous est donnée par le livre de Jean-Pierre Changeux, *L'Homme neuronal*. Cet ouvrage se présente comme une défense des mérites de la démarche réductionniste en neurobiologie, mérites d'autant plus spectaculaires que le niveau ultime de réduction descend dans l'échelle de la matière[39] . Or, lorsqu'il s'agit de fonder la thèse philosophique que « l'homme n'a dès lors plus rien à faire de l'“Esprit”, il lui suffit d'être un Homme Neuronal[40] », c'est exactement au niveau où Warren McCulloch l'a placée que l'argumentation se déroule, et dans les mêmes termes.

Une indication plus directe nous est fournie par l'attitude de McCulloch et Pitts, relayés par Wiener, lors de la discussion (passionnante pour l'histoire de la neurophysiologie) qui ouvrit la sixième conférence Macy (1949). Von Neumann, absent, avait transmis un message : « 1010 neurones, traités comme des relais simples, sont tout à fait insuffisants pour rendre compte des capacités humaines », spécialement en matière de mémoire. Il suggérait que les degrés de liberté (ou de complexité) manquants se trouveraient dans l'organisation interne, protéinique, du soma, le corps cellulaire du neurone. Les physiologistes présents abondèrent dans ce sens, trop heureux de dénoncer le simplisme du neurone axiomatisé de McCulloch. Ni celui-ci ni Wiener n'esquivèrent la discussion, très technique sur les plans biochimique et physiologique. Mais, à lire les Actes, on sent parfaitement que McCulloch, jamais péremptoire ni dogmatique dans les discussions, ne voyait pas en quoi cela pouvait affecter la pertinence de son modèle. Le débat fut clos lorsque Pitts démontra que le calcul sur lequel von Neumann avait fondé sa critique était invalide[41].

Une comparaison vient à l'esprit, qui est d'autant moins hors de propos que la *Théorie des jeux* de von Neumann et Morgenstern, achevée en cette même année 1943 et publiée l'année suivante, va fournir un ingrédient appréciable au débat cybernétique. C'est la représentation de l'acteur par la théorie économique. Un économiste pur et dur peut s'intéresser à titre personnel aux connaissances accumulées par la psychologie dans son exploration des profondeurs de l'âme humaine, cela ne l'incitera en rien à renoncer **à** son modèle de l'*homo oeconomicus*, ce calculateur « input-output » qui est à l'homme complet ce que le neurone formel est au neurone réel. Il y a une profonde communauté d'esprit entre la modélisation cybernétique et la modélisation en économie mathématique, et l'on ne s'étonne donc pas que les nombreux avatars de la première (théorie des systèmes, recherche opérationnelle, théorie du contrôle optimal, théorie de la décision, etc.) aient procuré à la seconde maints de ses outils.

Il est donc absurde de dire ou d'écrire, comme on l'entend ou le voit trop souvent, que le modèle de neurones idéalisés ne relève que d'une neurologie « formelle » et « fictive ». Autant dire, remarque McCulloch, que la théorie de l'électromagnétisme de Maxwell (ou n'importe quelle théorie physique formalisée) est fictive parce qu'elle se donne des êtres idéalisés. Il est de fait que la communauté neurobiologique est pour l'esssentiel toujours restée indifférente à l'œuvre de McCulloch. Même ceux qui reprennent l'idée centrale que la caractéristique déterminante du système nerveux est d'être un réseau de neurones, et utilisent des formalismes issus du modèle de McCulloch et Pitts, semblent ignorer l'existence de ce modèle, ou bien l'attribuent à d'autres auteurs[42] . Très rare exception, le neurobiologiste Jerome Lettvin (qui fut, il est vrai, l'un des collaborateurs du grand cybernéticien) considère que Warren McCulloch a édifié ce qui, aujourd'hui encore, est la seule *théorie* du système nerveux dont nous disposions, et attribue le dédain, l'ignorance ou l'ostracisme dont elle fait l'objet au manque d'ambition théorique et philosophique de la discipline. Les données anatomiques et physiologiques s'accumulent mais on ne songe guère à les mettre en rapport avec les processus mentaux[43].

Ironie ou ruse de l'histoire, ce n'est ni en logique, ni en neurobiologie que le modèle de McCulloch et Pitts aura engendré une postérité notable, c'est en intelligence artificielle. Ironie, car de tous les cybernéticiens, McCulloch fut le plus éloigné des préoccupations de l'ingénieur. Ruse – mais pour le montrer, il faudrait contempler toute l'histoire des sciences cognitives, de la cybernétique à aujourd'hui : cette histoire reste encore à écrire[44], et il serait sans doute prématuré de le faire, tant les situations évoluent rapidement. Je me contenterai ici de quelques brèves indications.

Après la phase cybernétique, qui est le principal objet de ce livre, la recherche sur les réseaux neuronaux connut un essor appréciable qui culmina dans les travaux que Frank Rosenblatt réalisa entre 1957 et 1961 à l'université Cornell, pour le compte de la Marine américaine qui s'intéressait à la modélisation de la perception visuelle. Le modèle de Rosenblatt, qui connut son heure de gloire sous le nom de « Perceptron », était un réseau de McCulloch et Pitts particulier, arrangé en couches de neurones superposées et capable d'apprentissage[45]. Les neurones d'entrée simulent l'activité de la rétine, les neurones de sortie classent les traits que le système « reconnaît » dans l'objet qu'on lui présente, et les neurones des couches cachées réalisent les calculs intermédiaires. Dans le Perceptron, seules les connexions entre la couche de sortie et la dernière couche de calcul étaient capables d'apprentissage.

Dans l'ouvrage de 1962 qu'il a consacré à ses recherches, Rosenblatt définit ce qui distingue sa démarche de celle de McCulloch et Pitts. Ceux-ci partent des performances dont ils veulent que leur réseau de neurones soit capable, et ils en déduisent la structure de ce dernier. C'est une méthode constructive, descendante (*top-down*, comme on dit en anglais), qui implique que l'organisation logique du réseau est définie de façon rigoureuse, sans raté ni gaspillage. Première ironie, Rosenblatt lui reproche d'être bien peu biologique. A cette démarche qu'il nomme « monotypique », il oppose une méthode « génotypique », consistant à partir d'une classe de réseaux dont la structure est définie de manière incomplète, floue, statistique, et à en explorer les propriétés fonctionnelles (démarche *bottom-up*)[46]. Tout en admettant sa dette envers McCulloch et Pitts, Rosenblatt se reconnaît comme influences plus immédiates celles du cybernéticien de la seconde phase, Ross Ashby et… de l'économiste et philosophe social Friedrich von Hayek. Nous y reviendrons. Même si c'est de manière un peu injuste que le concepteur du Perceptron

prend ainsi ses distances par rapport à McCulloch, sa critique marque un tournant dans l'évolution du sens à donner aux réseaux neuronaux. Il s'agit déjà moins d'une modélisation logique, et déjà plus de l'exploration des ressources d'un réseau complexe d'éléments simples en interaction.

Or, à l'époque, l'intelligence artificielle naissante s'engageait résolument sur d'autres voies, comme nous l'avons vu au chapitre précédent. Alors que la cybernétique dans sa seconde phase conservait l'ambition de modéliser l'intelligence naturelle et maintenait le contact avec la neurologie, l'intelligence artificielle se délivrait de cette contrainte et liait son sort au développement des ordinateurs. Si l'une comme l'autre assimilaient la pensée au calcul, la première continuait de situer les opérateurs de ce calcul au niveau neuronal, tandis que la seconde leur fixait pour domaine un niveau psychologique dont elle postulait l'autonomie, celui des représentations mentales – ce qui allait constituer la marque distinctive de la science cognitive orthodoxe, ou « cognitivisme ». Les champions de l'intelligence artificielle n'allaient pas manquer de déclarer la guerre au Perceptron, sans doute excédés par la publicité de mauvais goût qui entourait ce dernier, alors qu'ils estimaient que sa formalisation mathématique manquait par trop de rigueur. Le coup de grâce fut donné par Marvin Minsky et Seymour Papert en 1969, dans leur ouvrage *Perceptions*[47]. Ils prétendaient porter un coup d'arrêt aux fols espoirs qui avaient été mis dans ce type de modélisation en en démontrant les limitations formelles. A vrai dire, ces limitations condamnaient seulement les modèles les plus simples – ainsi, un réseau à deux couches se révèle incapable d'« apprendre » le connecteur logique « ou exclusif » (soit la fonction booléenne XOR) –, mais les auteurs ne craignaient pas de sonner le glas de la modélisation par réseau en général, arguant des difficultés apparemment insurmontables qu'il y avait à la complexifier.

Dans les années soixante et soixante-dix, la cybernétique, devenue « seconde cybernétique » ou « cybernétique du second ordre », se survit à elle-même et se montre capable de développer un courant de recherches sur les « systèmes à auto-organisation ». Les chercheurs en intelligence artificielle tiennent généralement ces travaux pour quantité négligeable. Dans un autre livre de 1969, intitulé *Semantic Information Processing*[48] , Marvin Minsky se livre à une analyse historique dans laquelle il évalue ce qu'aura été la fécondité de la cybernétique. Il présente celle-ci comme un tronc commun qui se serait divisé en trois branches : la « simulation cognitive » à la Newell et Simon ; l'intelligence artificielle proprement dite ; et enfin, une « branche morte », la seconde cybernétique ou théorie des systèmes auto-organisateurs. Dix ans plus tard, Hubert Dreyfus, le farouche pourfendeur de l'intelligence artificielle, s'autorise de l'évaluation de Minsky, pourtant très sérieusement éreinté par ailleurs, pour se dispenser de considérer le cas de la cybernétique, jugeant qu'elle n'a produit aucun résultat digne d'intérêt[49].

Connexionnisme *vs* cognitivisme

Il y a une justice immanente. Le début des années quatre-vingt aura été témoin d'un renouveau très puissant des recherches sur les réseaux de neurones, dorénavant nommés « neurones formels » pour bien marquer qu'il ne s'agit surtout pas de marcher sur les brisées de la neurobiologie. L'occasion en a été la découverte de nouvelles techniques d'apprentissage impliquant les diverses couches de neurones cachés, et surtout

l'application à ces réseaux de méthodes de traitement issues de la physique des systèmes désordonnés, et plus spécialement de la théorie des « verres de spin » – laquelle traite du comportement collectif des atomes aimantés d'un cristal lorsqu'il est soumis à un champ magnétique extérieur. La richesse des propriétés que ces techniques, qui s'inspirent parfois de la thermodynamique, révèlent dans les réseaux neuronaux est sans commune mesure avec ce qu'on pouvait en dire dans les années quarante et cinquante. Ces propriétés peuvent s'interpréter en termes cognitifs : les réseaux sont capables d'apprendre, de reconnaître des formes, de mémoriser par association, etc. Tant et si bien que ce courant de recherches se pose aujourd'hui en principal rival de l'intelligene artificielle orthodoxe, laquelle s'est entre-temps considérablement essoufflée, comme si elle avait atteint son seuil de compétence.

Une bonne partie des débats qui agitent les sciences cognitives à présent relève de l'affrontement entre les deux modèles, promus au rang de paradigmes : un cognitivisme orthodoxe, pour qui penser, c'est calculer comme un ordinateur, c'est-à-dire sur des symboles qui ont à la fois une réalité matérielle et une valeur sémantique de représentation ; le paradigme des réseaux neuronaux, pour qui penser, c'est calculer comme le font ces réseaux, de façon massivement parallèle, les comportements intéressants n'apparaissant qu'au niveau collectif, en « émergeant » du système des interactions entre computeurs élémentaires simples.

Toute la question aujourd'hui porte sur l'articulation entre les deux paradigmes. Daniel Memmi distingue quatre positions ou attitudes possibles[50] : la « ségrégation », qui attribue aux réseaux et au cognitivisme des domaines de pertinence disjoints, les premiers s'appliquant surtout à la reconnaissance des formes et aux mémoires associatives, alors que le second modéliserait les formes « supérieures » de la cognition (par exemple, le raisonnement et le langage naturel) ; la « compilation » (au sens que ce mot a en informatique), qui relègue les réseaux au rôle subalterne de « réaliser » au niveau le plus bas les modèles cognitivistes, de la même façon que, dans un ordinateur, les instructions données en langage de haut niveau sont traduites dans un langage que la machine peut « comprendre » ; l'« hybridation », qui consiste à concevoir des systèmes cognitifs alliant et articulant les performances de l'un et de l'autre modèle, ce qui implique que l'on sache comment traduire l'un dans l'autre ; et enfin, l'« inclusion », qui donne le rôle principal, voire unique, aux réseaux neuronaux, toute la difficulté étant de montrer que les règles logiques de calcul sur représentations, que l'intelligence artificielle orthodoxe ou le cognitivisme se donnent au départ, émergent comme régularités ou quasi-régularités du fontionnement du réseau ; de telle sorte que le paradigme des réseaux neuronaux inclurait le cognitivisme comme cas particulier limite, de la même façon, si l'on veut, que la relativité généralisée inclut la gravitation newtonnienne comme cas limite, pour des vitesses faibles devant celle de la lumière.

Nous n'avons pas encore, à dessein, rappelé le nom sous lequel le modèle des réseaux neuronaux est aujourd'hui connu : le « connexionnsime ». L'expression, à vrai dire, fut déjà utilisée par Rosenblatt[51] en 1958. Pendant un certain temps, on a dit « néo-connexionnisme ». Personne ne semblant savoir à quel « archéo » (ou « paléo »[52]) ce « néo » se rapportait, on l'a supprimé. La référence au modèle de McCulloch et Pitts était ainsi perdue à jamais. Une science dynamique n'a pas le temps de se retourner sur son passé.

Ce qui aurait pu passer pour la revanche posthume de McCulloch ressemble plutôt en

fin de compte à son second enterrement. Revenons sur les quatre positions théoriques dégagées par Daniel Memmi : ce sont évidemment les deux dernières qui donnent le beau rôle au connexionnisme. Or l'une comme l'autre impliquent que celui-ci se rapproche fortement du cognitivisme, même si c'est finalement pour mieux l'avaler. Les connexionnistes d'aujourd'hui posent donc, à propos de leurs réseaux, des questions qui n'étaient pas du tout celles de McCulloch – la première d'entre elles étant celle de la représentation symbolique du « monde extérieur ». Nous y reviendrons au chapitre suivant.

La machine de von Neumann

Revenons aux années quarante, qui sont notre propos. Nous avons annoncé trois moments. Le premier assimile l'esprit à une machine logique ; le deuxième fait de même avec le cerveau, et puisqu'il s'agit de la même machine, le cerveau et l'esprit ne font qu'un. Reste le troisième moment. C'est seulement en ce dernier temps, et non au premier comme le veut la légende, que la machine artificielle, matérielle, à savoir l'ordinateur, entre en scène. Les faits, pourtant, sont bien avérés. Les dates, d'abord, sont éloquentes. C'est seulement en 1943 – décidément, une année propice – qu'est mise sur le chantier la construction de la première calculatrice électronique ultra-rapide, l'ENIAC, à la Moore School of Electrical Engineering de l'université de Pennsylvanie. Si étonnant que cela puisse nous paraître aujourd'hui, ce dinosaure informatique a été réalisé sans que l'on disposât de l'idée que la conception logique d'une machine à calculer est *séparable* de la conception de ses circuits, telle qu'elle se trouve contrainte par l'état de la technique et la nature des composants physiques – ou, pour le dire en termes actuels, sans que fût faite la distinction entre le « matériel » (le *hardware*) et le « logiciel » (le *software*). Cette idée, c'est John von Neumann qui la formule alors que, appelé en consultation par les constructeurs de l'ENIAC, il élabore déjà les concepts de la prochaine génération d'ordinateurs. Or d'où la tire-t-il, sinon de la lecture qu'il fait de l'article de McCulloch et Pitts ? Face à ce nouveau venu dans le monde matériel qu'est l'ordinateur, von Neumann adopte exactement la même attitude que ces derniers face au cerveau biologique : il en *abstrait* une machine logique-et, qui plus est, la *même* machine logique. On dit parfois plaisamment que si Turing fut le père de l'ordinateur, von Neumann en fut le médecin accoucheur ou la sage-femme. Il manque de toute évidence à cette paire un élément indispensable : la matrice. Disons que ce fut la machine de McCulloch.

Sans ce travail logique sur la machine artificielle, l'idée constitutive de l'ordinateur moderne n'aurait probablement pas été réalisée dans la matière. Cette idée, c'est que l'on peut mettre au même niveau les règles de fonctionnement, les instructions – on dira plus tard le « programme » – et ce sur quoi elles portent : les données. C'est cette confusion maîtrisée de ce que la logique élémentaire demande de distinguer : l'opérateur et l'opérande, qui fait de la machine matérielle une machine universelle au sens de Turing. Déjà celle-ci réalisait la même mise à plat, puisque sur son ruban (sa « mémoire ») on juxtapose le code de la machine particulière qu'il s'agit de simuler, c'est-à-dire les instructions, et le code des données. Il n'est nul besoin de mémoires séparées pour ces deux ensembles d'information. Cependant, pour « remonter », comme le fit von Neumann, de la machine matérielle à la machine de Turing, il fallait un intermédiaire, et

ce fut la machine de McCulloch.

L'ENIAC était une machine parallèle, comme celle de McCulloch. La machine conçue par von Neumann allait s'appeler l'EDVAC et fonctionner de façon séquentielle. C'est en comparant automates naturels et automates artificiels que von Neumann devait théoriser cette innovation. Les neurones sont extrêmement nombreux dans le cerveau et ils fonctionnent lentement : c'est l'organisation en parallèle qui optimise le traitement de l'information. En revanche, les composants d'un calculateur artificiel peuvent atteindre de très grandes vitesses et ils sont relativement peu nombreux : il faut les monter en série. Aujourd'hui encore, presque tous les ordinateurs – on les appelle parfois « machines de von Neumann » – sont des machines séquentielles. Cette domination est si forte que même les réseaux neuronaux du connexionnisme, ces machines hautement parallèles, sont *simulés* sur des machines de von Neumann. Ce n'est que l'un des chassés-croisés dont cette histoire n'est pas avare[53].

Norbert Wiener, Arturo Rosenblueth, Julian Bigelow, Warren McCulloch et Walter Pitts : les auteurs des deux articles fondateurs parus en 1943 sont ceux qu'à l'instar de Steve Heims nous nommerons les « cybernéticiens »[54] . Mais, de même que les trois Mousquetaires étaient quatre, les cinq cybernéticiens étaient six. Du d'Artagnan de l'équipe, certes plus juif d'Europe centrale que gascon, mais doté de la même faconde et rayonnant du même panache, nous avons par la force des choses déjà beaucoup parlé. Il n'est cependant pas possible d'enfermer la personnalité de John von Neumann à l'intérieur de la seule cybernétique. Ses travaux, qui ont, pour le meilleur ou pour le pire, ouvert à la science et aux techniques des chemins nouveaux dans un nombre impressionnant de domaines, recoupent cette dernière en deux points au moins. C'est d'abord sa *Theory of Games and Economie Behavior*, publiée en 1944 en collaboration avec l'économiste autrichien Oskar Morgenstern. C'est surtout son élaboration d'une théorie générale des automates, naturels et artificiels, dont nous avons déjà dit qu'elle avait trouvé son impulsion initiale dans la lecture de l'article de McCulloch et Pitts. Nous venons de rappeler le rôle décisif que von Neumann allait jouer dans la conception de ces automates artificiels que sont les ordinateurs actuels. On connaît moins, en général, sa contribution au problème même qui tourmentait McCulloch : les rapports circulaires entre logique et cerveau. Von Neumann dut apprendre beaucoup de neurophysiologie pour pouvoir élaborer une comparaison systématique du cerveau et de l'ordinateur. Il en présenta les prolégomènes lors de sa conférence historique au symposium Hixon, en 1948, sous le titre « The General and Logical Theory of Automata ». Plus tard, il allait aborder deux problèmes qu'il jugeait essentiels pour la compréhension de la logique des organismes : comment des composants non fiables peuvent constituer un automate fiable ; et quels principes d'organisation sont suffisants pour donner à un automate la capacité d'autoreproduction. Sa mort prématurée, en 1957, devait interrompre ces recherches. Mais en 1948, tout en rendant hommage à la découverte de McCulloch et Pitts, il prend ses distances par rapport à elle, de façon fort intéressante[55] . Il commence par évoquer la question de la taille du cerveau capable d'incarner telle ou telle fonction, ou propriété de l'esprit. Comme nous avons déjà eu l'occasion de le voir, aux conférences Macy et ailleurs, il conjecturera que le nombre de neurones formels nécessaires à la production de ces propriétés risque fort d'être très nettement supérieur au nombre de neurones réels. Puis vient sa critique essentielle. Tout comportement descriptible sans ambiguïté en un nombre fini de mots est computable par un réseau de neurones formels, c'est là un résultat

remarquable, reconnaît-il – trop rapidement, pour la raison que nous avons expliquée. Mais nos comportements les plus complexes – et par « comportements », von Neumann entend ici, en bon cybernéticien, les facultés de l'esprit : pensée, formation des concepts, mémoire, apprentissage, etc. – sont-ils descriptibles, en pratique, totalement et sans ambiguïté, au moyen de mots ? Localement, c'est toujours possible, par exemple s'il s'agit de décrire notre capacité à reconnaître la même forme triangulaire derrière deux triangles empiriques, de tracés, de tailles et de positions différents. Mais s'il s'agit de caractériser globalement notre capacité à établir des « analogies visuelles » ? Alors, conjecture von Neumann, on peut être dans un cas où le moyen le plus simple de décrire un comportement est de décrire la structure qui l'engendre – le gain en complexité de description pouvant tendre vers l'infini. Il est, dans ces conditions, dépourvu de sens de « découvrir » que tel comportement peut être incarné dans un réseau de neurones puisqu'il n'est pas possible de définir le comportement autrement qu'en décrivant le réseau. La morale de l'affaire, von Neumann ne fait que l'esquisser, car il est assez paradoxal pour lui, mathématicien, de la défendre face au neurophysiologiste McCulloch : celui-ci expliquait la neurophysiologie au moyen de la logique ; il est plus fécond, suggère von Neumann, d'enrichir notre logique sur le modèle de la neurophysiologie. Il répétera souvent que la logique formelle, dans son état présent, est trop rigide, trop combinatoire pour pouvoir se poser comme *la* logique des automates, qu'ils soient artificiels ou naturels ; il cherchera lui-même à édifier une logique plus complexe, plus proche du continu, s'inspirant dans ses méthodes de la théorie des probabilités, de la thermodynamique, de la théorie de l'information et de l'analyse mathématique[56].

Von Neumann posait ainsi la question de la *complexité*, en prévoyant qu'elle deviendrait *la* grande question de la science à venir. La complexité, cela impliquait pour lui la futilité de la démarche constructive à la McCulloch et Pitts, qui réduit une fonction à une structure ; cela imposait à la place la question de savoir *de quoi* une structure complexe est capable. Il est significatif que Marvin Minsky, qui passait sa thèse sous la direction de von Neumann, ait jugé que la critique de son patron à l'encontre de la démarche de McCulloch était une aberration, un aveu de faiblesse, un manque de foi dans ce que lui-même, John von Neumann, avait réussi à édifier[57].

Il reste à se demander si cette critique assez radicale eut une fécondité sur le mouvement cybernétique. Nous savons que McCulloch en fut ébranlé et qu'il consacra les dernières années de sa vie à des recherches sur des logiques alternatives, mais sans beaucoup de succès, semble-t-il. Ross Ashby, Frank Rosenblatt et, plus tard, le néoconnexionnisme allaient, en sens inverse de l'intelligence artificielle orthodoxe, poser la question dans les termes de von Neumann, se donnant une classe de structures définies seulement en partie et de façon statistique, et cherchant à en explorer les possibilités générales ou « génériques ». On note cependant avec amusement que dans la carte qu'il dresse en 1984 des sciences cognitives, Dan Dennett range le modèle de McCulloch du même côté que l'orthodoxie cognitiviste, l'autre côté étant représenté par le « *néo*-connexionnisme »[58]. Comme si l'incarnation par celui-ci du paradigme de la complexité pouvait justifier non seulement de couper le lien de paternité avec la cybernétique de McCulloch, mais encore de confondre celle-ci avec ce qui, dans l'ensemble des sciences cognitives, lui ressemble le moins : le cognitivisme orthodoxe. Comme si la cybernétique était vouée à jamais à se voir rejetée par ses enfants.

3

Les limites de l'interdisciplinarité

> « Je crois qu'avec la cybernétique les hommes ont mordu au fruit défendu comme ils ne l'ont jamais fait depuis deux mille ans. »
>
> Gregory BATESON[1]

L'histoire que nous racontons est essentiellement une histoire intellectuelle. Quelques faits et dates sont cependant nécessaires pour fixer les idées. En mai 1942 fut tenue à New York une conférence sur l'inhibition dans le système nerveux central, sous les auspices d'une fondation médicale philanthropique, la fondation Josiah Macy Jr. Deux amis de longue date s'y retrouvèrent, deux physiologistes, Arturo Rosenblueth et Warren McCulloch. Il y avait là aussi, outre le directeur médical de la fondation, Frank Fremont-Smith, Lawrence Frank, une sorte d'entrepreneur intellectuel touche-à-tout, ex-administrateur de la fondation, le psychanalyste Lawrence Kubie et le couple d'anthropologues Gregory Bateson et Margaret Mead[2] . Il y fut beaucoup question d'hypnose, mais Rosenblueth présenta aussi les idées de ce qui allait devenir le fameux article de 1943[3] . Bateson dira plus tard le choc intellectuel que ce fut pour lui. McCulloch, sentant les résonances entre ces idées et celles qu'il développait par ailleurs avec Pitts, proposa à Fremont-Smith d'organiser un cycle de conférences autour de ce qui ne s'appelait pas encore cybernétique[4] , sur le modèle de ceux que la fondation tenait par ailleurs sur des thèmes médico-sociaux très variés qui pouvaient aller de la régulation de la pression artérielle aux problèmes de l'enfance, en passant par les maladies du foie, le métabolisme, la coagulation sanguine, l'influx nerveux et les problèmes du vieillissement[5]. Fremont-Smith accueillit la suggestion avec intérêt. La chose devait se faire, mais seulement après la guerre.

Après cette rencontre, les échanges allaient s'intensifier entre le groupe de McCulloch à Chicago et celui de Wiener au MIT. Le rapprochement se concrétisa en 1943 par la venue de Pitts au MIT, auprès de Wiener, avec la bénédiction de McCulloch. Pour sa part, McCulloch espérait attirer Rosenblueth à Chicago. Celui-ci, qui venait de perdre son directeur de laboratoire à Harvard, Walter Cannon, préféra retourner dans son pays, où il monta un laboratoire de physiologie au sein de l'institut national de cardiologie de Mexico. Pitts et Wiener devaient dès lors parcourir très souvent le triangle Cambridge (Mass.)-Mexico-Chicago.

Parallèlement, Wiener poursuivait des échanges passionnés avec von Neumann sur la question des analogies entre organismes et machines. Ils en étaient au point de penser que les choses étaient mûres pour une institutionnalisation des recherches dans ce domaine, sous la forme d'une revue, d'une société savante, et même d'un centre de recherche. Une rencontre fut organisée en janvier 1945 à l'Institute of Advanced Studies de Princeton, où

travaillait von Neumann[6] . Il y avait là, entre autres, von Neumann et Wiener, Pitts, Goldstine, qui participait à la conception de l'ENIAC, McCulloch et Lorente de No'[7] . Wiener allait en sortir plus que jamais convaincu que « l'engineering et la neurologie ne font qu'un » et qu'un « programme de recherche permanent » devait être mis sur pied incessamment[8].

Le MIT était le lieu idéal pour le centre rêvé. Wiener s'arrangea pour qu'il fît une offre intéressante à von Neumann. Celui-ci se servit de cette offre pour faire monter ses enchères personnelles à Princeton. L'Institut était très réticent pour laisser von Neumann y construire son *computer* ultra-rapide : il se laissa convaincre. Bien que déçu, Wiener n'en voulut pas à von Neumann, et il lui envoya même son ingénieur préféré, Bigelow[9]. Voilà comment la première cybernétique n'a jamais eu de « centre » bien à elle. Les conférences Macy devaient en quelque sorte y suppléer.

Les conférences Macy

La guerre finie, la fondation Macy reprit la suggestion que lui avait faite McCulloch en 1942 et chargea celui-ci de monter un cycle de conférences sur les idées naissantes, qui n'avaient pas encore reçu de nom de baptême. Selon les principes de la fondation, il s'agissait de réunir à intervalles réguliers (en général, tous les six mois) un petit groupe d'une vingtaine de chercheurs, membres officiels du cycle en question, lesquels pouvaient s'adjoindre à l'occasion jusqu'à cinq « invités ». L'accent était mis beaucoup plus sur les échanges et les discussions que sur les exposés formels. La première rencontre eut lieu en mars 1946 à New York. Le titre en était « Feedback Mechanisms and Circular Causal Systems in Biological and Social Systems »-ce qui pouvait passer pour une référence aux travaux de Wiener, c'est évident, mais aussi à ceux de McCulloch, étant donné l'importance que les circuits fermés de neurones avaient dans sa théorie. L'inclusion des « systèmes sociaux » dans le programme résultait de l'initiative de Bateson. Ce dernier, convaincu dès 1942 de la pertinence des idées nouvelles pour les sciences sociales, aida McCulloch à choisir les participants dans le domaine des disciplines « molles ». Parmi ceux-ci, le plus fort contingent était celui des psychologues : gestaltistes comme Mollie Harrower, sociaux comme Kurt Lewin, expérimentalistes comme Heinrich Klüver, analytiques comme Lawrence Kubie[10] . Le sociologue Paul Lazarsfeld et le mathématicien Leonard J. Savage étaient également présents[11] ; ainsi, bien sûr, que les anthropologues Bateson et Mead.

L'automne 1946 fut une période d'intense activité pour le groupe. En septembre, reprenant une suggestion faite par Lazarsfeld lors de la première conférence, Bateson se chargea d'organiser hors cycle, mais toujours sous les auspices de la fondation Macy, une conférence spéciale destinée aux chercheurs en sciences sociales, pour leur permettre de dialoguer avec Wiener et von Neumann. Le titre en était « Teleological Mechanisms in Society »[12] . Les sociologues Talcott Parsons et Robert Merton étaient parmi les participants, ainsi que l'anthropologue Clyde Kluckhohn[13] . En octobre, eut lieu la deuxième conférence Macy du cycle animé par McCulloch, sous le nouveau titre de « Teleological Mechanisms and Circular Causal Systems »[14] , puis, immédiatement après, un symposium organisé par Lawrence Frank pour l'Académie des sciences de New York, sur le thème : « Teleological Mechanisms ». Frank, qui savait, selon ses propres termes,

comment créer un « climat d'opinion », présenta en termes lyriques les idées du groupe, affirmant que leur enjeu était « la construction d'un nouveau cadre de référence conceptuel pour la recherche scientifique dans les sciences de la vie[15] ». Ce fut à cette rencontre, d'esprit nettement plus « militant » que les conférences Macy, que Norbert Wiener exposa à un important public de scientifiques, probablement pour la première fois, une vision unifiée de la théorie, centrée sur les notions – à la terminologie désormais fixée – de message, communication, information, feedback, automate, etc. Ce sont *ses* vues, beaucoup plus que celles de McCulloch, qu'il développa, passant sans beaucoup de précautions de la machine à l'organisme et de celui-ci à la société[16]. Il revint à McCulloch de clore le symposium en présentant une « récapitulation de la théorie » des mécanismes téléologiques et en ouvrant quelques pistes de recherche. Habileté, conviction ou soumission, il moula l'exposé substantiel de ses propres idées dans le cadre général fixé par Wiener.

En 1947 fut organisée la troisième conférence Macy sous le même titre que la deuxième[17] . C'est également l'année où Wiener forge le mot de « cybernétique » pour donner une unité au mouvement d'idées[18] . En 1948, paraît son ouvrage du même nom, et sont tenues les deux dernières conférences du premier cycle, intitulées « Circular Causal and Feedback Mechanisms in Biological and Social Systems »[19]. 1948 est aussi l'année d'un important symposium organisé au California Institute of Technology (CalTech) par le Hixon Fund Committee. Lors de ce symposium Hixon, von Neumann et McCulloch sont confrontés à la critique des plus grands noms de la neurophysiologie (Karl Lashley, Ralph Gerard), de la psychologie (Wolfgang Köhler) et de l'embryologie (Paul Weiss) de l'époque, d'une façon beaucoup plus vive, voire houleuse, qu'au sein des conférences Macy, sur lesquelles McCulloch exerce son influence et sa maîtrise.

En février 1949 débarque aux États-Unis, venant de Vienne, un physicien autrichien spécialisé dans l'ingénierie électrique, Heinz von Foerster. Il a dans sa poche une monographie sur la mémoire, dans laquelle il recourt aux concepts de la mécanique quantique pour modéliser le caractère de « tout ou rien » de l'influx nerveux. Ce texte parvient entre les mains de McCulloch, lequel invite aussitôt von Foerster à participer à la sixième conférence Macy, la première du nouveau cycle. L'anglais de von Foerster étant des plus sommaires, on décide, pour lui faire apprendre très vite la langue de Jefferson, de le nommer secrétaire du cycle de conférences, chargé d'en établir les comptes rendus. C'est ainsi que von Foerster fut d'emblée plongé dans le bain cybernétique[20]. C'est lui, on le sait, qui deviendra le chef de file de la seconde cybernétique. Sa première action fut de faire approuver par le groupe unanime la décision d'intituler désormais les conférences Macy : « Cybernetics ». Le sous-titre de « Circular Causal and Feedback Mechanisms in Biological and Social Systems » fut néanmoins maintenu.

De 1949 à 1953, une fois par an, à peu près à la même époque, furent tenues les cinq dernières conférences Macy sur la cybernétique, les seules dont les Actes ont été publiés. Dès la huitième, en 1951, Wiener et von Neumann sont absents. Dès la neuvième, leurs noms disparaissent de la liste des « membres » du cycle. L'histoire de cette période se termine tristement. En 1952, Pitts convainquit McCulloch de venir le rejoindre au MIT. Peu de temps après, pour des raisons mal élucidées, mais où se mêlent des rivalités personnelles et un scandale familial, Wiener rompit toutes relations avec le groupe. Ce divorce devait marquer profondément McCulloch, Pitts et leurs amis pour le restant de leur vie[21].

Le matériau dont dispose l'historien des idées est d'abord constitué des Actes des cinq dernières conférences Macy. Comme le note von Foerster, le fait que ces Actes soient devenus depuis longtemps très difficiles à consulter a puissamment contribué à donner aux conférences Macy le caractère sacré d'un mythe fondateur. Quant aux cinq premières conférences, il n'en reste aucun enregistrement. Cependant, nous disposons d'une lettre que McCulloch adressa à l'avance à tous les participants de la quatrième conférence « sur les mécanismes téléologiques » (1947), qui se présente comme un compte rendu détaillé des trois premières[22] . Les Actes de la dixième conférence contiennent également un résumé des neuf premières[23], établi par McCulloch. Les Actes de la huitième et de la neuvième conférences sont précédés d'une « Note by the Editors », qui donne des indications utiles sur l'histoire du cycle. On trouve dans l'« Introduction » de Wiener à *Cybernetics* (1948) quelques données sur les toutes premières conférences.

Le travail patient d'historien réalisé par Steve Heims depuis de nombreuses années – lequel a comporté interviews des protagonistes encore en vie, dépouillement des archives personnelles des uns et des autres[24] , analyse de leur correspondance, etc. – lui a permis de nous donner une vision plus exacte des cinq premières conférences, certainement les plus importantes pour la naissance de la cybernétique. C'est ainsi, entre autres résultats remarquables, qu'il a réussi à reconstituer l'affrontement entre le gestaltiste Köhler et le groupe des « cybernéticiens » qui marqua la quatrième conférence, en 1947[25] . Cette rencontre qui devait illustrer les possibilités de synergie entre la psychologie gestaltiste et la démarche physico-mathématique fut un échec, et cet échec retentit sur la teneur des conférences ultérieures. Quant à la cinquième conférence, dont nous savons[26] qu'elle fut consacrée à la structure du langage, avec la participation de Roman Jakobson, il en existe un bref compte rendu.

Ces indications sont nécessaires, car la dissymétrie de notre information entre la première et la seconde parties du cycle des conférences sur la cybernétique introduit un biais dans notre appréciation de l'ensemble. Il est possible que ce biais soit important, surtout pour qui veut comparer l'influence de McCulloch à celle de Wiener – une comparaison dont nous allons voir qu'elle est décisive pour qui veut comprendre ce que fut le projet de la cybernétique. En effet, McCulloch, en tant que président, n'aura pas manqué une conférence, alors que Wiener disparaît dès la huitième. Les Actes publiés ne le mettent donc en scène que deux fois. Sans doute a-t-il joué, avec von Neumann, un rôle décisif dans la première partie du cycle, celle dont nous savons qu'elle fut la plus mouvementée et qu'elle servit à définir l'identité du groupe, sa composition et son programme de discussions. Mais d'un autre côté, lorsque Steve Heims pose la question, effectivement cruciale, de savoir ce qui a permis à un groupe interdisciplinaire, qu'aucune activité de recherche commune ne rassemblait sinon de se réunir et de discuter, de durer aussi longtemps[27] , et qu'il attribue cette stabilité à l'influence « dominante » de Wiener[28], n'est-ce pas là, une fois de plus, un effet de la vieille injustice commise à l'égard de McCulloch, qui fait que seul le nom de Wiener reste attaché au projet cybernétique, puisque, celui-ci disparu, les conférences continuent avec le même entrain ?

Quoi qu'il en soit, les Actes publiés des cinq dernières conférences représentent un document irremplaçable. Comme lors des rencontres elles-mêmes, dont ils nous rendent compte avec la plus extrême précision, l'essentiel en est constitué non par des communications formelles, mais par des discussions, souvent fort vives, des échanges rapides, des exposés courts et constamment interrompus, où se révèlent les caractères, les

oppositions, les intérêts et les coalitions. C'est bien ainsi que la fondation Macy concevait ces rencontres : comme des expériences en vraie grandeur, destinées d'une part à manifester ce qui n'apparaît jamais dans la présentation par un scientifique de ses résultats, à savoir les processus de décision, de sélection et d'exclusion de pistes de recherche qui y conduisent ; d'autre part, à illustrer les difficultés psychologiques, sémantiques et autres, qui entravent la communication entre spécialistes de disciplines différentes, et à suggérer des moyens de briser ces obstacles. En d'autres termes, les conférences Macy étaient déjà conçues comme devant fournir un matériau de première main à ce qu'on appellerait bien plus tard des recherches de type « Science-Technique-Société[29] ».

Nous avons joint à ce corpus les Actes du symposium de l'Académie des sciences de New York sur les « Teleological Mechanisms », ainsi que ceux du symposium Hixon, à titre de documents-témoins. A cette dernière rencontre, les cybernéticiens n'étaient pas en position de force, et il est intéressant de comparer leur maintien, en cette occasion, à la façon dont ils débattaient lors des conférences Macy. Autre élément précieux de comparaison : les Actes du symposium Alpbach de 1968, organisé par Arthur Koestler, sur le thème : « Beyond Reductionism[30] ». A ce grand moment de la pensée systémique participeront, entre autres, Paul Weiss, Ludwig von Bertalanffy, Friedrich Hayek et C. H. Waddington[31]. Même s'il n'était pas désigné ainsi, le thème de l'auto-organisation était omniprésent, et il servit d'arme pour une critique de la cybernétique. La comparaison est d'autant plus intéressante que l'embryologiste Paul Weiss, qui se trouvait là dans son milieu naturel, s'était opposé directement aux cybernéticiens, au symposium Hixon, quelque vingt ans plus tôt. Enfin, pas plus que nous ne l'avons fait jusqu'ici, nous ne feindrons, dans le regard que nous portons sur cette période, d'ignorer les événements subséquents, en particulier l'aventure de la seconde cybernétique et les débats actuels au sein des sciences cognitives.

Les « cybernéticiens » en discussion

Les conférences Macy virent un tel flux de participants et une telle diversité de thèmes de discussion que toute tentative d'y repérer de grandes masses est largement vouée à l'arbitraire. Nous prenons néanmoins le risque de proposer les répartitions suivantes, pour l'ensemble des dix conférences (ce qui ajoute aux chances d'erreur, étant donné l'insuffisance de l'information pour les cinq premières).

Pour ce qui est des participants, nous ne retenons que ceux dont les interventions nous ont semblé marquer d'une façon ou d'une autre un moment clé des débats[32]. Dans un premier groupe, nous mettons les six cybernéticiens, y ajoutons trois futurs éminents représentants de la seconde cybernétique : Heinz von Foerster, W. Ross Ashby et W. Grey-Walter, ainsi que les théoriciens de l'information, Claude Shannon, Donald MacKay, Yehoshua Bar-Hillel et Henry Quastler. Il n'est pas absurde de rajouter à ce groupe, qui comprend ainsi quatorze membres, le mathématicien Leonard Savage. Pendant toutes ces années, il mettait sur le chantier sa théorie de la décision, laquelle paraîtrait en 1954 sous le titre *Foundations of Statistics.* Malgré de sensibles divergences d'appréciation sur ce que peut la cybernétique naissante (Savage, en particulier, ne prenait pas au sérieux son ambition interdisciplinaire), tous ont en commun un certain rapport à la

science qui privilégie la modélisation mathématique.

Viennent ensuite les anatomistes, physiologistes, médecins et naturalistes, qui ont moins de goût pour l'abstraction que les précédents, et jouent le rôle de caution, de garde-fou ou de critique par rapport aux ardeurs modélisatrices de leurs collègues. Nous retenons huit noms : outre les grands neurophysiologistes Rafael Lorente de No' et Ralph Gerard, dont nous avons déjà parlé, il y a les médecins Frank Fremont-Smith et Harold Abramson, le chirurgien William Livingston, les anatomistes Gerhardt von Bonin et John Z. Young, et le zoologiste George E. Hutchinson.

Les psychologues forment à eux seuls une troupe impressionnante : nous en recensons quinze, plus que de « mathématiciens ». Ils représentent le continent à conquérir, soit qu'ils se laissent complaisamment faire et même, parfois, précèdent le mouvement ; soit qu'ils résistent vaillamment, voire défient les conquérants. Il y a parmi eux des gestaltistes comme Wolfgang Köhler et Mollie Harrower, et des psychologues sociaux d'obédience gestaltiste comme Kurt Lewin et son ancien élève Alex Bavelas ; des expérimentalistes : Heinrich Klüver, John Stroud, Joseph Licklider, Heinz Werner et Donald Marquis ; des psychiatres ou neuropsychiatres : Henry Brosin (qui présidera le symposium Hixon), Hans Lukas Teuber, Eilhardt von Domarus, auxquels il faut ajouter le psychanalyste Lawrence Kubie ; et enfin des comparatistes, spécialistes du comportement animal : Theodore Schneirla et Herbert Birch.

Vient enfin le petit groupe des représentants de la philosophie et des sciences sociales : les anthropologues Gregory Bateson et Margaret Mead, le sociologue Paul Lazarsfeld, l'éclectique Lawrence Frank et le philosophe Filmer Northrop ; auxquels il faut adjoindre le linguiste Roman Jakobson.

Dans sa sécheresse comptable, ce tableau suffit à nous donner une indication précieuse. Contrairement à une certaine image que la cybernétique a donnée d'elle-même, ce n'est pas tant aux sciences de la *vie* qu'elle s'est affrontée, pour puiser en elles son inspiration, pour les révolutionner ou les conquérir. Cette image a sans aucun doute été forgée par la pratique systématique et tous azimuts des analogies entre organismes et machines à laquelle Wiener et ses disciples se livraient. Non, l'interlocuteur premier, c'était les sciences de l'*esprit*. Cela est d'autant plus remarquable qu'à la même époque certains physiciens, à la suite de Niels Bohr, portent leur regard sur la biologie. On connaît le *What is Life* ? de Schrödinger, qui date de 1944, mais il y a aussi les travaux de Max Delbrück (qui fut l'élève de Bohr) sur les bactériophages, organismes très primitifs mais capables de se reproduire. Von Neumann, plus « *inside outsider* » que jamais[33] , s'était intéressé à ces recherches, et il avait mis Wiener au défi de lui prouver qu'il ne serait pas plus fructueux pour la pensée mathématique de modéliser le bactériophage plutôt que le cerveau[34] . Max Delbrück fut donc invité à devenir membre régulier des conférences Macy. Il participa à la cinquième, et jura qu'on ne l'y reprendrait plus. Le premier jour était, on l'a déjà dit, consacré au langage, avec comme invités Roman Jakobson, Charles Morris et Dorothy Lee. Le lendemain, Wiener fit une analyse des mécanismes susceptibles d'incarner le « démon » de Maxwell et parla de la possibilité que l'ordre puisse surgir du chaos – thème qui devait, bien plus tard, connaître une certaine fortune, avec la seconde cybernétique et la physique des systèmes désordonnés. Pitts, lui, proposa une analogie formelle, bien dans un certain style « cybernétique », entre les collisions de molécules et l'établissement d'un ordre hiérarchique chez les poulets (le fameux *pecking order*). Max Delbrück devait déclarer plus tard que toute cette rencontre fut, de son point

de vue, « vide au plus haut point et complètement inepte[35] ». Jugement peut-être injustement sévère, mais que l'on peut interpréter ainsi. Avec Delbrück, c'était la physique, dans toute sa rigueur mais aussi toute son orthodoxie, fût-elle celle de la mécanique quantique, qui s'attaquait à la biologie. Schrödinger, dans son livre, avait en quelque sorte donné le feu vert, en prédisant que les mécanismes de l'hérédité et de la génétique seraient un jour réduits aux lois connues de la physique. Or il n'y avait aucun physicien professionnel chez les cybernéticiens. Certes, nous l'avons déjà dit et nous y reviendrons en y insistant fortement, car ce point est trop souvent ignoré, c'est au nom de la physique qu'ils menaient leur combat. Mais c'était une physique pour le moins singulière car, en une enjambée périlleuse, elle avait sauté par-dessus le vivant pour rejoindre directement la logique et l'esprit. Un physicien strict comme Delbrück y voyait surtout du charlatanisme.

L'ironie de l'histoire des sciences est grande, ici comme souvent. Quelques années plus tard, le groupe de Delbrück – qui s'était donné pour nom le « groupe du phage » – allait inventer la biologie moléculaire en découvrant le code génétique. Peut-on dire qu'était ainsi couronnée de succès l'ambition proclamée de réduire la biologie à la physique ? On sait que la biologie moléculaire a dû recourir à une terminologie fort étrange et bien peu… physicaliste pour décrire les mécanismes biochimiques responsables de la « téléonomie » de la vie – la « téléonomie », notion popularisée par l'équipe française qui participa à cette aventure, se voulant une version encore plus éloignée de toute finalité *intentionnelle* que la « téléologie ». Programme, code, information, transcription, message, traduction : tout ce vocabulaire de la biologie moléculaire est directement importé de la description des communications entre êtres humains. C'est donc à une physique repeinte aux couleurs de la psychologie qu'a été réduite la biologie ! En réalité, on sait qu'il y a eu un intermédiaire, et que cet intermédiaire est… la cybernétique, cette discipline qui, précisément, s'était donné pour objectif d'« effondrer la muraille qui sépare le monde magnifique de la physique du ghetto de l'esprit », pour reprendre les termes lyriques de Warren McCulloch[36]. La cybernétique semble avoir été condamnée à ne prendre de revanches que posthumes.

Il n'y eut donc pas de biophysiciens ni de biochimistes aux conférences Macy. Cette absence constitue en tout cas un nouveau signe que l'influence de Warren McCulloch – celui qui pensait l'incarnation de l'esprit dans la machine – n'était pas moindre que celle de Norbert Wiener – celui qui pensait l'organisme et la machine comme modèles l'un de l'autre. Une étude sommaire de la répartition en grands thèmes des échanges qui constituèrent l'ordre du jour des conférences Macy va dans le même sens. Nous avons défini de façon très approximative une « unité de discussion », laquelle coïncide en général avec une communication plus ou moins longue suivie d'un débat : mais il y a pas mal d'exceptions, car les échanges prenaient parfois des directions non prévues au programme. Nous comptabilisons ainsi dix-sept unités de discussion sur le modèle de la machine logique commune – ou non, c'est là l'objet du débat – au cerveau et à l'ordinateur, dont sept unités de discussions plus spécifiquement anatomiques et physiologiques, et quatre relatives au problème des *Gestalten*.

Pour ce qui est des analogies entre organismes et machines « à la Wiener », notre compte fait apparaître sept unités seulement, dont deux qui traitent de psychologie expérimentale. Il est cependant légitime de leur adjoindre quatre unités spécifiquement consacrées à la théorie de l'information, ce qui donne au total onze unités qui portent

plutôt la marque du style de Wiener – contre dix-sept qui relèveraient plutôt des préoccupations de McCulloch et de von Neumann (mais nous savons combien ce partage est arbitraire, et contraire aux objectifs déclarés des conférences Macy).

Plus surprenant, peut-être, est le faible nombre de discussions qui portent sur les « machines cybernétiques », rat de Shannon et autre homéostat d'Ashby, qui ont façonné l'image de la nouvelle discipline : en comptant large, nous trouvons cinq unités. Il est vrai que ces séances annoncent plutôt la seconde cybernétique.

Vient ensuite l'important groupe des discussions sur les névroses et la pathologie de la vie psychique : sept unités ; sur les communications humaines et sociales : onze unités, dont six spécifiquement sur le langage et une sur la théorie des jeux ; sur les communications anormales : deux unités ; soit vingt unités au total sur le thème de prédilection de la cybernétique : la communication, traitée comme comportement.

Il faut enfin noter la très faible part accordée aux discussions épistémologiques générales : on ne compte guère qu'une unité, consacrée à une question qui fera les délices de la seconde cybernétique mais à laquelle la première répond d'une façon infiniment plus positiviste : la place de l'observateur. Nous sommes à la sixième conférence. La discussion réunit Kubie, Wiener, Fremont-Smith et Stroud[37] . Kubie a posé le problème de l'observateur en psychanalyse : il doit être aussi « inhumain » et « détaché » que possible. Wiener, aussitôt, pense à une analogie physique : lorsque la lumière produit un effet de résonance sur la particule qu'elle permet d'observer. Stroud rêve à des expériences de psychologie où l'observateur serait une machine. Mead et Bateson protestent à peine. Il y a une étrange amnésie de la part de von Foerster à déclarer, dans un compte rendu tardif des conférences Macy, que celles-ci firent progressivement entrer l'observateur dans le royaume de la science dont il avait jusqu'alors été exclu[38] . C'est confondre ce qu'allaient être les préoccupations de la seconde cybernétique avec celles de la première. Von Foerster lui-même jouera un rôle décisif dans cette identification de la cybernétique et de l'épistémologie et dans l'édification d'une « épistémologie cybernétique ». Avec lui, la cybernétique allait, dans ses propres termes, devenir une cybernétique des systèmes observants, et non plus des systèmes observés[39]. C'est le trait que l'on retient aujourd'hui le plus souvent de la cybernétique, et l'école néo-cybernétique française n'est pas pour rien dans sa diffusion. Ce serait, cependant, une grave erreur d'appréciation que de le projeter sur ce que fut l'esprit de la première cybernétique. Il s'agissait plutôt alors d'atteindre, dans le domaine des sciences de l'esprit, au même degré d'objectivité qu'en physique.

Unifier le travail de l'esprit

On voit souvent aujourd'hui dans la cybernétique une tentative de renouveler les modes de pensée propres aux sciences de la nature, l'ambition de créer une « science nouvelle » qui prenne pour objet les relations entre éléments, et non les éléments eux-mêmes, une science des totalités engendrées par ces relations ; donc, une approche holiste, s'opposant au réductionnisme de la science « orthodoxe » ; un effort désespéré pour aller à contre-courant des tendances apparemment irréversibles de la science contemporaine, de plus en plus professionnalisée et spécialisée, et pour donner tort à l'avertissement terrible de Max Weber dans son essai de 1919, « La science comme vocation » : « En science, une

réussite incontestable ne peut être aujourd'hui qu'une réussite dans un domaine spécialisé. Et celui qui est incapable de porter des œillères, celui-là ferait mieux de se tenir à l'écart de la science[40]. »

Cette image s'est formée sur les avatars de la cybernétique : seconde cybernétique, théorie des systèmes ou systémique, théories de l'auto-organisation, etc. Gregory Bateson a largement contribué à la répandre dans son œuvre propre, et l'on peut soutenir avec Steve Heims qu'elle n'est pas contradictoire avec la personnalité et l'œuvre de Wiener[41] . Selon Heims, il pouvait y avoir des divergences philosophiques sérieuses entre Wiener, d'un côté, Rosenblueth et Bigelow, de l'autre, à propos de l'interprétation à donner de leur article commun de 1943. Le positivisme qui éclate à de multiples reprises dans les propos que Rosenblueth tient au cours des conférences Macy – nous en avons déjà vu un exemple – n'aurait pas du tout été partagé par Wiener[42]. Quoi qu'il en soit, ceux qui, aujourd'hui encore, brandissent cette image d'une cybernétique conquérante sont soit les champions purs et durs, mais attardés, d'une cybernétique marginalisée, soit, plus fréquemment, ses critiques féroces pour qui le simple rappel de ses ambitions suffit à la condamner sans appel. La question qui nous intéresse est : cette image correspond-elle à la réalité de la première cybernétique, telle qu'elle se manifeste lors des conférences Macy ?

Au début de chaque conférence, l'administrateur de la fondation, le docteur Fremont-Smith, ne manquait jamais de rappeler la vocation interdisciplinaire de ces rencontres. La nature ne connaît pas de frontières, martelait-il, aussi bien la spécialisation, la professionnalisation et l'isolement croissants des disciplines scientifiques constituent-ils aujourd'hui les principaux obstacles au progrès de la connaissance. D'où le mot d'ordre : abattre les cloisons artificielles, mettre les diverses spécialités en relation de communication afin de permettre une réunification de la science[43] . La *communication* était donc non seulement l'objet des conférences, mais aussi leur mode et leur raison d'être. Que cet effort fût indissociable d'une volonté de rupture épistémologique avec la « science orthodoxe » et d'une mise en accusation de la « méthode analytique », cela apparaît aussi dans les propos de certains des participants, au premier rang desquels Lawrence Frank. Celui-ci, dans son allocution d'ouverture au colloque de l'Académie des sciences de New York sur les « mécanismes téléologiques », ne craint pas de faire en bloc le procès du mécanisme ancien, de la « causalité linéaire » (qui manifeste la croyance animiste en des forces mystérieuses), du réductionnisme d'une science qui ne voit que les états et ignore les processus. En un œcuménisme hardi, il rapproche le néomécanisme cybernétique des idées nouvelles en physique. Citant Einstein et le concept de champ physique, Langmuir et sa notion de phénomènes « divergents », qui anticipe le concept de « sensibilité aux conditions initiales », lequel sera promis à un grand avenir avec la théorie du « chaos » – un événement minime en un point d'un système suffit à altérer sensiblement le tout qu'il constitue –, Schrödinger et le principe d'indétermination, et bien sûr les principales idées de la cybernétique, il estime que l'on dispose dorénavant de tous les ingrédients nécessaires à une théorie des totalités organisées. Et cela, finalement, … pour la plus grande gloire de la biologie : car, rappelle-t-il, voici ce que disait le grand biologiste Haldane, quelque soixante ans plus tôt, à propos de la physique de son temps : « Qu'un point de rencontre entre la biologie et la physique puisse être trouvé dans l'avenir, il n'y a aucune raison d'en douter. Mais nous pouvons prédire en toute certitude que si ce point de rencontre est trouvé, et que l'une des deux sciences doit être engloutie

par l'autre, cette science-là ne sera *pas* la biologie[44]. »

Le lecteur des Actes des conférences Macy qui s'en tiendrait à la rhétorique d'un Fremont-Smith, d'un Lawrence Frank ou d'une Margaret Mead, trois des participants les plus dynamiques, les plus présents et les plus militants, serait vite entraîné sur une fausse piste. Pour éventer le piège, il convient de replonger le cycle des conférences dans le contexte social et politique de l'époque, comme l'a fait Steve Heims dans son ouvrage *The Cyhernetics Group*. L'Amérique des années d'après guerre est traumatisée par les folies collectives qui viennent tragiquement de déchirer le monde. Un concept prend de l'importance et devient la clé qui permettra, croit-on, d'ouvrir la porte d'un nouvel ordre mondial enfin rendu à la paix : la « santé mentale ». La fondation Macy finance des organisations qui se créent ou se développent pour promouvoir cette valeur. Fremont-Smith, Frank et Mead aident à constituer la Fédération mondiale pour la santé mentale en 1948 ; le Comité international sur l'hygiène mentale, qui s'était déjà réuni avant la guerre, accueille dans son conseil d'administration plusieurs des participants aux conférences Macy : Fremont-Smith, Mead, Brosin, Harrower et Kluckhohn. Lawrence Frank résume parfaitement l'esprit de ces efforts lorsqu'il décrit la société de son temps comme une « patiente » que seule la science peut guérir.

La science, oui, mais quelle science ? Le credo de la Fédération mondiale pour la santé mentale nous donne une première indication : « Les sciences de l'homme nous donnent l'espoir de traiter d'une façon nouvelle les problèmes de la guerre et de la communauté mondiale… C'est le but ultime de la santé mentale que d'aider les hommes à vivre avec leurs semblables à l'intérieur d'un même monde… Le concept de santé mentale est coextensif à l'ordre international et à la communauté mondiale qui doivent être développés afin que les hommes puissent vivre en paix les uns avec les autres[45]. »

C'est donc la foi dans le pouvoir curatif, libérateur et pacificateur des sciences de l'homme qui anime nos pionniers. Mais il faut être plus précis. *La* science de l'homme qui recueille leurs faveurs est la *psychiatrie, conçue comme science sociale.* Divers courants ou écoles de pensée contribuent à délimiter un nouveau champ d'investigations, connu sous le nom de « personnalité et culture » : l'anthropologie culturaliste d'un Franz Boas, celui qui, avec Ruth Benedict, fut le mentor de Margaret Mead à l'université Columbia ; l'école de Chicago en sciences sociales, avec John Dewey et George Herbert Mead ; l'école psychiatrique de Harry Stack Sullivan et l'anthropologie sociale d'Edward Sapir, qui avait été lui-même l'élève de Franz Boas. La notion clé qui se dégage de ces divers travaux est celle d'une *causalité circulaire* entre personnalité et culture : la personnalité des individus est façonnée par le milieu social et culturel dans lequel ils vivent, mais inversement, ce milieu est le reflet de la personnalité de base de ceux qui l'habitent. Il devrait donc être possible d'agir sur la psychologie des individus afin d'induire les changements voulus dans la société globale, à condition de tenir compte des effets en retour qu'une telle intervention ne peut manquer de produire. On saisit que la place était déjà bien aménagée pour accueillir les idées nouvelles de la cybernétique, de la théorie de l'information et de la théorie des jeux, lesquelles apportaient la caution scientifique et technique indispensable[46].

On aura une bonne idée du zèle militant avec lequel ceux des participants aux conférences Macy qui se reconnaissaient dans le mouvement « personnalité et culture » entendaient diffuser le nouveau credo, en se reportant aux paroles inaugurales de Lawrence Frank à la conférence de l'Académie des sciences de New York sur les

« mécanismes téléologiques », à laquelle nous nous sommes déjà référés :

« On peut voir dans le concept de mécanismes téléologiques une tentative d'échapper aux formulations mécanistes anciennes, qui maintenant apparaissent inadéquates, et de fournir des conceptions neuves et plus fructueuses, ainsi que des méthodologies plus efficaces, pour étudier les processus autorégulés, les systèmes et organismes auto-orientés, et les personnalités autodirigées [...]. Il est loisible de tenir cette conférence pour une étape importante, et peut-être même majeure, dans la progression vers ce nouveau climat d'opinion qui est maintenant en train d'émerger dans les activités scientifiques, philosophiques et même artistiques. Nous ne sommes pas de simples témoins : par ces rencontres et discussions, nous sommes les artisans actifs de ce nouveau climat d'opinion [...]. La façon dont je vois les choses, c'est que nous sommes engagés, aujourd'hui, dans l'une des transitions ou renversements majeurs de l'histoire des idées [...]. Quand les sciences sociales accepteront les conceptions nouvelles [...] et apprendront à penser en termes de processus circulaires, alors elles feront probablement des avancées spectaculaires [...]. La nouvelle démarche a déjà fait ses preuves dans le domaine psychoculturel, où elle commence à éclairer le fait que les régularités socio-culturelles et les traits spécifiques de la personnalité sont des aspects duaux d'une même réalité[47]. »

Le docteur Frank Fremont-Smith, l'administrateur de la fondation, n'était pas en reste. Champion d'une interdisciplinarité qui allierait la physique à la psychologie, il soulignait l'urgence de trouver un langage commun à ces deux disciplines afin d'assurer la paix mondiale. Les physiciens, explique-t-il lors de la sixième conférence, en mars 1949, sont les spécialistes des armes de l'hostilité ; les psychologues, et singulièrement les psychanalystes, sont les spécialistes des motivations de l'hostilité – l'assemblée constituante de la Fédération mondiale pour la santé mentale n'avait-elle pas, l'année précédente, pris comme devise, l'empruntant à la toute récente constitution de l'UNESCO : « Puisque les guerres commencent dans l'esprit des hommes, c'est dans l'esprit des hommes qu'il faut construire la défense de la paix » ? Il faut donc que physiciens et psychologues collaborent dans le but de mettre les principes de la science et de la logique au service de la résolution des « problèmes du comportement collectif et de la paix mondiale ». Cela ne sera possible que si les psychanalystes restent au contact des sciences fondamentales, et si les physiciens, chimistes, ingénieurs et autres mathématiciens acceptent de se soumettre à l'initiation psychanalytique[48].

Fremont-Smith voyait une autre raison à cette collaboration entre sciences exactes et psychanalyse. Il s'agit des mécanismes de l'invention scientifique. La logique est nécessaire pour ordonner les connaissances et pour concevoir les tests qui les mettent à l'épreuve. Mais il n'y aurait pas de dynamique scientifique sans créativité ; or les processus de celle-ci relèvent de l'autre face de l'esprit humain, cette zone d'ombre où se tapissent l'inconscient et l'irrationnel. On ne peut donc avoir une vue d'ensemble du travail de l'esprit qu'en unifiant les compétences du mathématicien et celles du psychanalyste[49] . Margaret Mead ira souvent dans le même sens, en affirmant que c'est une infirmité de notre culture que d'avoir séparé et isolé ce qui ailleurs reste unifié : notre part obscure et notre part de lumière, la « grammaire des rêves » et ce genre de pensée rationnelle qu'un Walter Pitts juge seule « normale »[50].

La tentation physicaliste

A ce stade, nous devons faire une pause, et prendre du recul par rapport à notre propre exposé. Dans les deux premiers chapitres de ce livre, nous avons tenté de reconstituer ce que fut la mise en place d'une science physicaliste de l'esprit axée sur une nouvelle conception de la machine. Cette mise en place passait nécessairement par une confrontation des sciences mathématiques et physiques et des sciences traditionnelles de l'esprit, à savoir diverses sortes de psychologies. C'est bien ainsi que les cybernéticiens l'endendaient. La biologie elle-même n'était pertinente que dans la mesure où elle était déjà une biologie de l'esprit, mais non pas, nous y avons insisté, en tant qu'elle posait la question : « Qu'est-ce que la vie ? » Si les cybernéticiens avaient besoin de se confronter à des psychologues, c'était donc pour des raisons fort éloignées de celles qui motivaient leur hôte, Fremont-Smith. Leurs rapports à la psychologie étaient nettement plus prosaïques : ils y voyaient surtout un terrain à conquérir, une opposition à réduire ; éventuellement, une source d'informations à recueillir et de défis à relever. Wiener avait *son* psychologue, Edwin Boring, qui lui fournissait des listes de fonctions psychologiques considérées d'un point de vue behavioriste, la charge revenant à Wiener d'en concevoir les « équivalents » électriques ou électroniques – « équivalents » voulant dire ici capables des mêmes transformations d'input en output[51] . De même Heinrich Klüver, dès la première conférence Macy, mit McCulloch et Pitts au défi de rendre compte au moyen de leur réseau de neurones idéalisés de la capacité du cerveau à percevoir et à reconnaître des *Gestalten*[52] – défi que les deux cybernéticiens ne mirent guère de temps à relever[53] . Paradoxalement, le combat le plus long à mener fut celui qui paraissait le plus facile : le siège de la notion d'inconscient, remise inlassablement sur le tapis de presque toutes les conférences par le psychanalyste Kubie. Notion parfaitement inconsistante selon les cybernéticiens. Ou un événement se produit, ou il ne se produit pas, explique Rosenblueth à la huitième conférence. Ou un train d'impulsions nerveuses a été enregistré, ou il ne l'a pas été. Dire qu'il l'a été, mais « inconsciemment », est un non-sens[54]. Les cybernéticiens ne semblaient guère prêts à avoir avec la psychanalyse le type d'échanges réciproques que souhaitait un Fremont-Smith. Il s'agissait plutôt pour eux d'un obstacle à balayer.

La confrontation entre cybernéticiens et psychologues, qui était la raison d'être des conférences Macy, a reposé sur un énorme malentendu. Les promoteurs du mouvement « personnalité et culture » souhaitaient instaurer une réciprocité entre sciences mathématiques et physiques et sciences psychologiques établies (psychanalyse, psychologie du développement, *Gestalt*, phénoménologie et même behaviorisme) ; les cybernéticiens entendaient mener un combat contre ces dernières, au nom des premières. Loin de se confondre avec la recherche d'une synthèse générale tous azimuts, l'effort d'interdisciplinarité de la cybernétique était très précisément focalisé – et, pour faire pièce au mot de Haldane cité par Lawrence Frank, si une science risquait d'être « engloutie » par la *scienza nuova* qu'elle menaçait de devenir, ce n'était certainement pas la physique.

Lorsqu'on relit les Actes des conférences avec ce malentendu de base en tête, on arrive à reconstituer, à presque cinquante ans de distance, les nombreux quiproquos, souvent comiques, qui ont émaillé les échanges. A la sixième conférence, le « dialogue » entre Wiener et le médecin Harold Abramson, collègue et ami de Fremont-Smith, a quelque chose de surréaliste[55] : le premier expose les mérites de l'analyse dimensionnelle pour l'unification des sciences pendant que le second, en contrepoint, introduit l'idée que les physiciens et les psychologues doivent communiquer pour assurer la paix mondiale, après que Fremont-Smith eut expliqué que l'unification de toutes les disciplines scientifiques

passait par l'édification d'une science de l'homme. Plus comique encore est la façon dont plusieurs des représentants des sciences de l'homme se livrent, en fait, en victimes consentantes à l'appétit féroce des cybernéticiens. Le malheureux Kubie – qui, il est vrai, était un psychanalyste d'un genre assez particulier, puisqu'il s'était arrêté à la première topique (conscient, préconscient, inconscient) et croyait, comme Freud, à l'enracinement biologique des concepts psychanalytiques, ayant d'abord été neurophysiologiste – fit de nombreuses fois preuve d'un remarquable esprit de conciliation, pour ne pas dire de soumission, en dépit de son obstination à défendre la psychanalyse.

C'est ainsi qu'en réponse à la charge de Rosenblueth que nous avons déjà évoquée, Kubie reconnaît que la « psychiatrie » doit être abordée comme une « science naturelle », « objectiviste », reposant sur des concepts « opérationnels », et que la technique de l'analyse devrait présenter les mêmes qualités de reproductibilité que les expériences des sciences exactes[56] . Sur quoi Rosenblueth renchérit : il n'y a rien dans le langage et la démarche des sciences de la nature qui les rendrait inadéquats à traiter des problèmes dont s'occupe la psychiatrie[57] . Quant à Pitts, il explose : « Si les méthodes du psychanalyste ne sont pas scientifiques, c'est son problème de les rendre telles, et ce n'est certainement pas à nous, scientifiques, de les juger aussi bonnes que les nôtres[58]. »

Même Fremont-Smith et Frank, les plus idéologues du mouvement « personnalité et culture », rendaient hommage, chaque fois qu'ils le pouvaient, à la *supériorité* des sciences mathématiques et physiques. A la sixième conférence, Fremont-Smith émet le vœu que « le savoir, l'ingéniosité et la sagesse extraordinaires » des mathématiciens, des physiciens et des ingénieurs soient mis au service de la résolution du « problème humain ». Il demande en conséquence aux sciences sociales de parler un langage qui fasse sens pour les physiciens et les ingénieurs[59] . Lawrence Frank ira encore plus loin dans l'expression de la soumission des sciences de l'homme aux « vraies » sciences. A la neuvième conférence, il illustre la « réciprocité » souhaitable entre les sciences sociales et les sciences du système nerveux dans les termes suivants. Les premières demandent aux secondes : « Les hypothèses que vous faites nous permettent-elles de les reprendre à notre compte ? » ; tandis que les secondes demandent aux premières : « Les hypothèses que vous faites dans les sciences sociales sont-elles compatibles avec le savoir que nous possédons au sujet du système nerveux ? » Si l'on comprend bien, c'est une logique du type « Pile, je gagne ; face, tu perds »[60] . Le neurophysiologiste Lashley pourra déclarer au symposium Hixon, en recueillant, semble-t-il, l'assentiment général : « Ce qui nous réunit ici, c'est la conviction que, je pense, nous partageons tous, qu'il est possible en dernière instance de décrire les phénomènes de l'esprit et du comportement au moyen des concepts des sciences mathématiques et physiques[61]. »

Les physiologistes, pour leur part, étaient traités comme des conseillers, des garants, des informateurs. Certains, cependant, ne s'accommodaient pas de cet emploi subalterne. C'était le cas du neurophysiologiste Ralph Gerard. A la septième conférence, celui-ci mit solennellement le groupe en garde contre les dangers d'une formalisation mathématique qui irait beaucoup plus vite que l'observation et l'expérimentation. Il est intéressant de noter qu'il s'exprima alors comme s'il avait affaire à une OPA de la physique sur la biologie. Cela apparaît nettement au moins à deux endroits : d'abord lorsque Gerard, au début de sa communication « Some of the problems concerning digital notions in the central nervous system », fait allusion aux jeunes physiciens qui se précipitent sur le terrain de chasse de la biologie ; puis dans sa réplique ironique à une charge de Bigelow,

lequel vient de manifester on ne peut plus clairement une volonté de mainmise *physicaliste* sur l'étude du système nerveux[62].

On peut juger qu'il avait une perception biaisée de la situation, puisque les cybernéticiens n'étaient pas des physiciens et que leur proie n'était pas la biologie ; il n'en est pas moins vrai qu'eux-mêmes ne voyaient finalement pas les choses différemment – en tout cas en ce qui concerne l'identité du conquérant.

Loin, donc, de se penser en rupture avec la science « orthodoxe », la cybernétique se voyait comme le front avancé de celle-ci, prêt à occuper les territoires supérieurs de la création. Répétons-le : le combat qu'elle menait ne l'était pas au nom d'une quelconque « nouvelle science », mais au nom de la mathématique et de la physique.

La vérité est que le projet cybernétique se voulait l'apothéose de la physique, et cette vérité, c'est encore McCulloch qui l'a le mieux exprimée. « Obligeons le physicien à rendre compte de sa propre existence, en tant qu'il fait partie du monde physique », dit-il au début de sa communication au symposium Hixon. « Il est juste pour lui de s'en tenir à ses propres règles et donc de montrer en termes de masse, d'énergie, d'espace et de temps comment il en est venu à créer la physique théorique. Il doit alors se transformer en neurophysiologiste (c'est ce que j'ai fait), mais ce faisant il se trouve obligé de répondre à la question de savoir si la physique théorique est quelque chose dont on peut discuter en termes de neurophysiologie (et c'est bien ce qui m'est arrivé). Répondre non, c'est rester un physicien pur et dur. Répondre oui, c'est se transformer en métaphysicien – c'est du moins ce que l'on me dit. Mais est-ce exact[63] ? » Il est clair que pour McCulloch, permettre à la physique de rendre compte d'elle-même n'était pas la faire sombrer dans la « métaphysique », mais l'aider à réaliser son chef-d'œuvre. Ce qui est évidemment, dans un autre sens du mot que celui que lui attribue McCulloch, le comble de la métaphysique. Cet autre sens, c'est, avec Heinz von Foerster, Ross Ashby, Gregory Bateson et leurs disciples, la seconde cybernétique qui s'en fera le champion, en affirmant qu'avec la cybernétique la science s'élève à la conscience de soi et devient *épistémologie*[64].

La mise en scène de cette alliance confuse entre un projet de rénovation de la société par la science et une entreprise de conquête scientifique pourrait faire croire que l'on nageait constamment dans l'idéologie, aux conférences Macy. En fait, l'essentiel des discussions portait sur des sujets très « pointus », qu'ils relèvent de la neurophysiologie ou de la psychologie expérimentale. Comme l'explique von Foerster dans ses notes introductives à la huitième conférence Macy, l'effort d'unification entrepris par les cybernéticiens ne se situe pas au niveau des solutions, mais à celui des *problèmes*. Certaines classes de problèmes, définies par une même structure logique, traversent les disciplines les plus variées. La cybernétique s'est édifiée principalement autour de deux de ces classes : les problèmes de communication, d'une part ; les problèmes posés par l'étude des mécanismes qui produisent eux-mêmes leur unité (*self-integrating mechanisms*), d'autre part[65]. Le fait de parler surtout de problèmes évitait au groupe les dérives idéologiques les plus fâcheuses Von Foerster, dans les mêmes notes, souligne ainsi que les cybernéticiens n'ont développé aucune langue de bois propre en dépit, à l'époque, de six années d'association. Le jargon cybernétique se limitait à une poignée de termes : numérique/analogique, feedback et servo-mécanisme, causalité circulaire. Von Foerster, toujours lui, observe que ces termes sont utilisés avec beaucoup plus de parcimonie par les membres du groupe que par ses partisans extérieurs.

Les cybernéticiens ne montraient aucun complexe à parler technique avec les

spécialistes les plus divers, examinant avec la même aisance les résultats d'une expérience psycho-acoustique et une théorie sur les conditions dans lesquelles les mots du langage acquièrent une signification individuelle propre. Les quelques « généralistes » présents, au premier rang desquels Gregory Bateson, étaient souvent perdus, frustrés, ce qui les conduisait parfois à supplier que l'on n'oubliât pas la vocation universelle de la cybernétique. L'examen attentif des Actes donne d'ailleurs cruellement à voir un Bateson qui n'est pas « dans le coup », ignorant les concepts logiques les plus élémentaires – il confond, par exemple, le sens ordinaire du mot « quantification » avec celui qu'il a en logique[66] -, se méprenant sur le sens de ce qui se joue lors de ces conférences[67]. Les nombreux admirateurs de son œuvre se consoleront en disant que sa lecture erronée du projet cybernétique lui a précisément permis de développer sans entraves scientistes sa propre pensée…

S'il y avait peu de production idéologique à destination du groupe lui-même, y en avait-il, dans l'esprit des cybernéticiens, en direction de l'extérieur ? La question fut abordée au début de la septième conférence. Une série d'articles à sensation venaient d'être publiés par les grands magazines sur les activités du groupe, d'où il ressortait que celui-ci ayant démontré que le cerveau est une machine, on pouvait désormais envisager de construire des machines intelligentes. Certains participants en profitèrent pour condamner la crédulité, la paresse intellectuelle et le faible Ql moyen du grand public, l'accusant paradoxalement de faire au groupe un accueil enthousiaste – un *national fad*, dira Ralph Gerard – dont il se serait bien passé. Dans ce débat, Gerard se montre le plus ironique et Stroud le plus virulent envers le public, Wiener étant le seul à témoigner quelque tolérance à l'égard des articles parus dans *Life*, *Time* et *Newsweek*. Il avoue ne pas juger totalement répréhensible l'utilisation que fait cette prose de l'adjectif « pensant » à propos des machines[68].

C'est au fond une situation très ordinaire que celle-là, où les scientifiques rendent les « autres » responsables du destin de leurs idées et de leurs découvertes. Autre lieu commun qui n'aura pas manqué aux conférences Macy, celui du mathématicien (en l'occurrence l'ingénieur Bigelow) qui reproche aux représentants des savoirs « mous » de se laisser prendre au piège de la fascination pour les mathématiques[69]…

4

Philosophie et cognition

« La cybernétique est la métaphysique de l'âge atomique. »

Martin HEIDEGGER (1976)

Ce qui tient ensemble aujourd'hui les multiples programmes de recherche que l'on regroupe sous le nom de « sciences cognitives », c'est le travail philosophique qui est fait à leur propos. Sans la philosophie « cognitive », il y aurait des travaux en psychologie, en linguistique, en neurobiologie, en intelligence artificielle – il n'y aurait pas de science de la cognition. C'est la philosophie qui réfléchit et systématise la ou les attitudes de base qui constituent le seul lien social à l'intérieur du domaine. L'existence d'un lien social n'implique aucunement qu'il y ait un paradigme unique – on a vu qu'il y en a au moins deux, le paradigme cognitiviste classique ou orthodoxe, et le connexionnisme. Mais le choc entre ces deux modèles est lui-même créateur de solidarités. Ceux qui s'affrontent dans les controverses qui ponctuent l'histoire du champ se reconnaissent finalement moins comme adversaires que comme membres d'une même famille élargie. Or l'arbitre qui discipline, règle et finalement juge ces affrontements, c'est le philosophe.

Il y a, ici comme ailleurs, beaucoup d'ironie dans cette histoire. Les sciences cognitives se présentent volontiers comme la reprise à nouveaux frais par la science des questions philosophiques les plus anciennes concernant l'esprit humain, son organisation, sa nature, les relations qu'il entretient avec l'organisme (le cerveau), avec autrui et avec le monde. Mais l'identité de ce qui se donne pour science de l'esprit reste profondément philosophique. Cette science qui parle au nom des sciences et des techniques qui composent le domaine (encore une fois, principalement les neurosciences, l'intelligence artificielle, la psychologie dite cognitive et la linguistique), et auxquelles elle apporte ce supplément d'âme (ou d'esprit) qui les réunit les unes aux autres, est en réalité une philosophie. C'est une philosophie qui se glisse à l'intérieur du cheval de Troie des sciences et des techniques pour investir le domaine de l'esprit, et en chasser les intrus qui occupaient encore la place : d'autres philosophies – principalement les philosophies de la conscience, la phénoménologie, l'existentialisme –, d'autres psychologies – comme le behaviorisme et la psychanalyse –, d'autres sciences – singulièrement les sciences sociales et les sciences de l'homme de type structuraliste.

Naturaliser l'épistémologie

Quelle est donc cette philosophie « cognitive » ? Le nom qu'elle se donne à elle-même est « philosophie de l'esprit » (*philosophy of mind*). C'est aujourd'hui la branche la plus active et la plus florissante de la philosophie analytique. Mais dire cela, c'est énoncer un

paradoxe, lequel reste en grande partie une énigme pour la philosophie analytique elle-même lorsqu'elle se penche sur son histoire.

Il est en effet bien connu que le geste inaugural accompli par les divers « pères » ou courants fondateurs que la philosophie analytique se reconnaît est une rupture avec le « psychologisme ». Notons d'emblée que ce geste est aussi celui qui fonde l'autre grande philosophie contemporaine, la phénoménologie. Ce parallèle n'est pas fortuit car, nous allons le voir, la philosophie de l'esprit partage avec la phénoménologie l'un de ses concepts de base. Que l'on songe au formalisme de Hilbert ou au logicisme de Frege, de Russell et de Carnap, pour qui il est essentiel de couper les ponts entre la logique et la psychologie, l'objectivité de celle-là devant se fonder sur autre chose que la contingence ou la facticité de celle-ci ; au positivisme logique du cercle de Vienne, qui conçoit la philosophie comme une activité d'épuration du langage de la science pour en chasser toute « métaphysique » ; à Wittgenstein et à la philosophie du langage ordinaire de l'école d'Oxford, pour qui il n'y a pas de « langage privé » et, donc, le seul accès aux pensées passe par l'analyse du langage en tant qu'il est une activité publique soumise à des normes reconnues par une communauté intersubjective : dans tous les cas, il y a à la fois refus déterminé de fonder la philosophie sur la psychologie, et priorité donnée à l'étude du langage. La philosophie analytique est, de par ses origines, une philosophie du langage.

Au *linguistic turn* du début du siècle aurait cependant succédé un *cognitive turn*. Le rejet du psychologisme aurait donc finalement accouché d'une philosophie cognitive qui se donne pour psychologie philosophique. Même si la philosophie analytique est beaucoup moins encline que sa consœur, qu'elle dit « continentale », à revenir sans cesse sur son propre passé, elle ne peut manquer de s'interroger sur ce singulier revirement. Les raisons qu'elle en donne sont diverses. La première tient à l'évolution interne de la philosophie du langage. Tant le programme chomskyen de la grammaire générative que le courant pragmatique issu des travaux de Paul Grice impliquent que l'on « entre dans la tête » des locuteurs et des interlocuteurs : sans appel à certaines capacités cognitives des sujets, de type combinatoire dans le premier cas, de type inférentiel dans le second, on ne saurait rendre compte, affirment ces nouveaux courants, de la nature et des propriétés du langage et de la communication verbale. Un deuxième type de raisons mobilise le développement et les progrès des sciences cognitives elles-mêmes. Désormais, grâce à elles, il est possible de mener à bien le programme que Quine appelait de ses vœux : la « naturalisation de l'épistémologie ». Les questions que posait traditionnellement la philosophie concernant les fondements de l'objectivité de nos connaissances sont dorénavant susceptibles de recevoir des réponses fondées sur les sciences empiriques. Ce sont des processus causaux, en principe réductibles aux lois de la physique, qui expliquent la formation de nos connaissances, en tant que croyances justifiées, et plus généralement le fait que nos « états mentaux » sont adaptés au monde extérieur[1].

Un instant de réflexion montre que ces raisons n'expliquent rien Les raisons que la philosophie avait de se couper de la psychologie, et plus généralement des sciences de la nature, pour mener son enquête sur la validité objective des connaissances et la *légitimité* même d'une science de la nature – cette enquête ouverte par Kant et qu'il nomma « transcendantale » –, ces raisons, donc, étaient principielles, et parfaitement indépendantes de l'état d'avancement des sciences empiriques. L'enquête transcendantale sur la teneur en vérité des connaissances n'est pas la limite d'une enquête factuelle sur la

genèse effective de ces mêmes connaissances lorsque la précision et la fiabilité de cette enquête tendent vers l'infini. Pour le dire autrement : le *quid juris* n'est pas la limite d'un *quid facti*, une question de droit ne peut recevoir le même type de réponse qu'une question de fait, le devoir-être n'est pas réductible à l'être.

Ces distinctions kantiennes étaient la raison même de l'antipsychologisme d'un Frege, comme d'ailleurs de l'antipsychologisme d'un Husserl. Justifier le tournant cognitif de la philosophie analytique par le progrès des sciences cognitives est donc parfaitement vain.

La vérité est peut-être que le changement manifesté par l'émergence d'une philosophie cognitive au sein de la philosophie analytique est beaucoup moins radical qu'il n'y paraît, et ce pour deux raisons : la première est que, même si elle ne le sait pas assez, la philosophie cognitive conserve quelque chose de l'héritage kantien, donc continue à prendre ses distances avec le « psychologisme » ; la seconde, c'est qu'elle ne s'est pas vraiment déprise de la priorité donnée au langage.

Le premier point a été soutenu, dans un article important, par Joëlle Proust, tout du moins en ce qui concerne ce qu'elle appelle la « philosophie implicite » de l'intelligence artificielle[2]. La façon dont les représentants attitrés de celle-ci, tels Allen Newell et Herbert Simon, se représentent leur travail, et surtout le présentent à leur public et à leurs commanditaires lorsqu'il s'agit de défendre et d'illustrer les ambitions d'une discipline nouvelle, ne saurait tromper, suggère Proust. Il s'agit alors de montrer que l'intelligence artificielle est une science respectable qui, bien que science de l'« artificiel », contribue à la connaissance empirique de la nature : construire une machine qui incarne une hypothèse sur la réalité, ou constitue un modèle de celle-ci, et mettre à l'épreuve cette hypothèse ou ce modèle en faisant fonctionner la machine, c'est être fidèle à la méthode expérimentale qui prévaut dans les sciences de la nature.

En vérité, montre Joëlle Proust, le souci même d'universalité manifesté par l'intelligence artificielle permet d'interpréter sa démarche sous un tout autre éclairage : l'intelligence artificielle est une philosophie, et une philosophie de type transcendantal. Elle est à la recherche des conditions formelles de l'activité cognitive qui sont communes à tous les systèmes capables d'une telle activité, qu'il s'agisse de sujets humains, d'animaux ou de machines. Elle explore tous les modes possibles d'intelligence, au-delà de ceux que l'homme est susceptible d'exhiber. L'objet de son enquête, ce sont finalement les conditions *a priori*, nécessaires et suffisantes, qui tout à la fois rendent la connaissance possible et fondent l'objectivité, c'est-à-dire l'universalité, de cette connaissance.

Certes, la solution proposée par l'intelligence artificielle à ce problème n'est pas la solution kantienne. Le sujet transcendantal y est remplacé par le « système symbolique physique » ; à l'universalité de la synthèse *a priori* répond l'universalité de la machine de Turing. Mais dans l'un et l'autre cas, la distinction entre psychologie et critique de la connaissance, entre lois contingentes de la cognition et règles nécessaires, est soigneusement marquée. On dira qu'en naturalisant et en mécanisant sa quête transcendantale, l'intelligence artificielle la prive par là même de son sens. Mais n'oublions pas que ce qui garantit l'universalité de la machine de Turing n'est qu'une thèse, de caractère métaphysique – au sens où Popper parle d'un « programme métaphysique de recherche ». La métaphysique kantienne, comme science des pouvoirs *a priori* du sujet de la connaissance, n'est pas si éloignée.

Je n'oublie pas que l'objet de mon enquête est la phase cybernétique de la genèse des

sciences cognitives. Ce détour par l'histoire de la philosophie de l'esprit n'est cependant pas inutile. Ce que Joëlle Proust dit de l'intelligence artificielle peut déjà être dit de ce qui la précède, et en particulier de la démarche de Warren McCulloch. On retrouve chez lui cette même tension entre le désir de faire accéder la science de la nature à l'étude de l'esprit et le souci de la quête philosophique. D'un côté, il y a son ambition de bâtir une « épistémologie expérimentale », dont le modèle de neurones idéalisés constitue le plus beau fleuron ; de l'autre, la référence à Kant est très présente, et McCulloch entend relever le défi consistant à donner une base physicaliste au jugement synthétique *a priori*. J'ai rappelé plus haut qu'il était le premier à souligner que les mécanismes *matériels* que le modèle neuronal représente étaient suffisants, mais non point nécessaires, pour rendre compte des facultés de l'esprit. Cela ne veut pas dire qu'il n'était pas, *au niveau formel*, à la recherche de conditions nécessaires et suffisantes. Seymour Papert, déjà cité, ancien disciple de McCulloch et pionnier lui-même de l'intelligence artificielle, nous assure du contraire, dans sa préface à *Embodiments of Mind*[3] . Commentant cette préface, Jean Mosconi écrit : « Il y a [dans la démarche de McCulloch] davantage que l'attitude banale que l'on pourrait rencontrer chez tout constructeur de "modèles" de l'activité nerveuse. A ceux qui se contentent de l'adéquation de "tests pragmatiques", McCulloch oppose, dit Papert, une "quête rationaliste du nécessaire et de l'intelligible" ("a rationalist quest for necessity and comprehension"), qui se manifeste en particulier par sa volonté d'édifier une construction théorique indifférente aux choix contingents que l'on peut faire dans la caractérisation des neurones formels. Qu'en fait l'étude se restreigne parfois à des cas particuliers n'enlève pas, selon Papert, à la généralité du projet ; c'est plutôt affaire de style ou effet d'un manque d'instruments mathématiques[4]. » A quoi l'on peut ajouter que McCulloch, fort de son théorème d'équivalence, était, comme les autres cybernéticiens, pénétré du caractère universel de la machine de Turing, donc de sa propre machine.

L'obstacle de l'intentionnalité

Revenons à la philosophie analytique. La seconde raison qui conduit à fortement relativiser l'importance de son « virage cognitif », c'est que la philosophie de l'esprit qui en a résulté reste profondément marquée par la philosophie du langage. C'est toute la question de la représentation qui se trouve ici posée.

Dans l'une des rares études qui confrontent ce qu'est devenu le connexionnisme aujourd'hui au modèle de McCulloch et Pitts, Daniel Andler reproche à ce dernier d'avoir totalement ignoré le problème de la représentation[5] . Andler défend la thèse, que nous avons déjà mentionnée au chapitre deux, selon laquelle l'atout ou la chance du connexionnisme contemporain est d'avoir acquis les moyens de se raccorder au courant cognitiviste orthodoxe (même si c'est pour mieux l'englober), et ce, en lui empruntant pour l'essentiel sa notion de représentation. Rien ne s'oppose en principe à ce que le réseau neuronal actuel, par exemple de type « PDP » (*parallel distributed processing*[6]), manifeste les propriétés cognitives supérieures qui étaient jusqu'ici l'apanage de l'intelligence artificielle classique, échappant ainsi au ghetto de la perception où les recherches de type Perceptron semblaient s'être enfermées. On peut en particulier espérer que le connexionnisme rendra bientôt compte de la capacité d'un réseau à « produire » de la logique, cette capacité étant une propriété *émergente* du réseau dans son ensemble – ce

qui signifie qu'elle ne peut être localisée à aucun niveau inférieur, et en particulier pas à celui de ses éléments. Comment cela ? a-t-on envie de protester, McCulloch n'a-t-il pas fourni depuis longtemps la démonstration que chaque neurone peut être interprété comme un calculateur logique, réalisant par exemple l'un des connecteurs élémentaires ; et que le réseau dans sa globalité peut être traité comme un calculateur symbolique universel ? Non, répond Andler avec détermination, sans notion de représentation, rien de tel ne peut être dit démontré. Que la logique ait servi à modéliser le réseau n'implique pas que le réseau puisse être considéré comme un modèle de la pensée logique. Ce qui manque, philosophiquement, à l'argument de McCulloch, c'est de montrer comment le signal envoyé par tel neurone peut représenter une proposition sur le monde ; comment le codage qui permet d'établir l'équivalence entre un contenu de pensée et un calcul opéré par le réseau, comment ce codage, donc, peut recevoir une signification en termes de représentation.

L'argumentation de Daniel Andler est forte et convaincante. Il est à noter, toutefois, qu'elle tient pour acquis le fait même que les sujets connaissants *ont* des représentations ; et que le mode d'être de la représentation correspond à l'idée que s'en fait la philosophie cognitive. Pour cette dernière, un « état mental » représente quelque chose dans la mesure où il a un contenu, et ce contenu porte sur le monde. La philosophie cognitive a recours à un terme technique pour désigner cette capacité représentative des états mentaux : ils sont dits « intentionnels ». L'intentionnalité des états mentaux, dit-on encore en anglais, c'est leur *aboutness* : le fait qu'il sont *à propos de* certains objets. Ces objets sont hétérogènes aux états mentaux, ce sont des choses en soi (ou des propriétés de ces choses, ou encore des relations entre plusieurs d'entre elles). On pourrait d'emblée objecter que la philosophie cognitive n'est pas ici en phase avec le mouvement général dans lequel s'incrivent les sciences cognitives, tel que je l'ai présenté au premier chapitre. Si penser, c'est simuler, alors, comme l'écrit Joëlle Proust, poursuivant sa lecture kantienne de l'intelligence artificielle à la Newell et Simon, « faire référence, pour un système symbolique physique, c'est nécessairement *simuler* symboliquement la structure et les propriétés d'un objet. La principale présupposition de la simulation est qu'elle préserve la structure de la réalité. Cependant la notion de "réalité" doit à son tour être comprise dans les termes du système […]. L'idée de donnée radicalement hétérogène au système "connaissant" n'a […] simplement pas de signification[7] ». En termes kantiens : l'objet de la représentation, ce n'est pas la chose en soi, c'est l'apparaître de celle-ci à l'intérieur même de la représentation sous la forme du phénomène acquérant une validité objective.

Si, en revanche, on pose que l'objet de la représentation, c'est la chose en soi, comme le fait la philosophie cognitive, alors ce ne peut certainement pas en constituer le *contenu*. Lorsque je pense que j'ai oublié de cadenasser mon vélo, celui-ci ne se trouve pas contenu, avec son métal et son cuir, dans mon esprit. La chose en soi n'existe pas dans la représentation, elle en est constitutivement absente. Quelle est donc la nature du contenu de la représentation ? C'est la réponse que la philosophie de l'esprit apporte à cette question qui continue à en faire une philosophie du langage, malgré qu'elle en ait. Cette réponse est linguistique, et ce à deux niveaux. Soit ces états mentaux qui, pendant longtemps, ont capté l'attention des sciences cognitives[8] et que l'on nomme depuis Bertrand Russell « attitudes propositionnelles ». Comme leur nom l'indique, ils sont censés relier une « attitude psychologique » du type « croire que », « désirer que », « craindre que », « avoir l'intention de », etc., à une proposition portant sur le monde.

Dans la version fonctionnaliste, computationnelle et représentationnelle de la philosophie de l'esprit telle qu'elle est défendue par un Jerry Fodor ou un Zenon Pylyshyn, cette proposition s'exprime dans une phrase du « langage de la pensée », langage privé dont les symboles s'inscrivent dans la matière du cerveau.

Cette hypothèse du langage de la pensée n'est certes pas partagée par tous les philosophes de l'esprit, loin de là ; mais tous, ou presque, admettent que le critère de l'« intentionnalité » du mental est linguistique, à un second niveau : les phrases de la langue publique à laquelle nous avons recours pour attribuer à autrui des états mentaux doués de contenu – cette langue étant de préférence l'anglais, un peu comme le grec et l'allemand étaient pour Heidegger les langues du voilement/dévoilement de l'Être –, ces phrases, donc, possèdent la propriété d'être intensionnelles. On entend par là qu'elles violent les règles de l'extensionnalité logique, la première de ces règles étant la généralisation existentielle. De la vérité de : « La vache de Maurice broute dans le pré », on infère qu'il existe nécessairement un pré dans lequel broute la vache de Maurice. En revanche, ni la vérité ni la fausseté de : « Maurice croit que les dahus sont plus gras en Savoie que dans les Dolomites » ne permettent de conclure à l'existence ou à l'inexistence des dahus.

La seconde règle violée par une phrase intensionnelle est la substituabilité de termes ayant la même référence. Mon fils, qui vient d'avoir douze ans, ne sait pas que Tegucigalpa est la capitale du Honduras ; en revanche, malgré son peu de goût pour la logique, il sait que la capitale du Honduras est la capitale du Honduras. C'est cette propriété que Willard Van Orman Quine a rendue fameuse sous le nom d'« opacité référentielle ». Mais c'est en fait un autre philosophe américain, Roderick Chisholm, qui, dans un livre publié en 1957 sous le titre *Perceiving*[9] , a le premier proposé cette interprétation linguistique de l'intentionnalité. Il prend l'exemple suivant : « La plupart d'entre nous savions en 1944 que Eisenhower était l'homme qui avait le commandement ; mais bien qu'il fût (identique à) l'homme qui allait succéder à Truman, il n'est pas vrai que nous savions en 1944 que l'homme qui allait succéder à Truman était celui qui avait le commandement[10]. »

Une grande partie des difficultés que rencontre la philosophie de l'esprit aujourd'hui provient de ce choix initial. Bon nombre de ses représentants acceptent la caractérisation linguistique de l'intentionnalité tout en s'efforçant de la « naturaliser » – c'est-à-dire d'en fournir une analyse fondée en dernière instance sur les lois de la physique. Le problème est qu'ils souhaitent conserver dans cette entreprise « physicaliste » quelque chose que la psychologie ordinaire tient pour acquis, à savoir que les contenus des états mentaux ont une pertinence causale dans l'explication de nos comportements. Si Maurice est allé en Savoie plutôt que dans les Dolomites chasser le dahu, c'est *parce qu*'il croyait *qu*'il y trouverait des animaux plus dodus. L'obstacle a paru jusqu'ici insurmontable, parce que la philosophie de l'esprit s'est convaincue que le contenu sémantique d'un état mental, apprécié par ses conditions de vérité et de référence, dépend de tout l'environnement physique et social du sujet ; or, si ce contenu a un pouvoir causal au sens de la physique, ce pouvoir ne semble pouvoir se concevoir qu'en termes des propriétés intrinsèques de l'état mental. Le prix à payer pour naturaliser la théorie de la connaissance paraît donc être de priver les propriétés et les faits mentaux de toute efficacité causale en tant qu'ils sont mentaux – en faisant ainsi de purs « épiphénomènes »[11].

L'une des configurations théoriques les plus originales que l'effort pour surmonter

l'obstacle a suscitées est le « monisme anomal » de Donald Davidson, exposé dans un article de 1970, « Mental Events[12] ». Davidson postule que tout *événement* mental est identique à un événement physique ; en revanche, il n'y a pas de relation d'identité entre classes, ou types, d'événements mentaux et classes, ou types, d'événements physiques. Les propriétés, mentales ou physiques, étant précisément des classes d'événements singuliers, cette position philosophique combine un monisme ontologique – « il n'y a », en dernière instance, que des événements physiques – et un dualisme des concepts et des propriétés : les concepts mentaux sont irréductibles aux concepts physiques. Ce monisme non réductionniste est « anomal » dans le sens suivant. Les événements mentaux causent d'autres événements mentaux et des événements physiques, mais ce que la *relation* de causalité relie, ce sont les événements en tant qu'événements du monde, indépendamment de la description, mentale ou physique, que nous en donnons. La relation causale est donc extensionnelle, et elle est sous-tendue par une loi de la physique. En revanche, une *explication* qui fait intervenir des propriétés ou des concepts mentaux, à l'instar des explications fournies par la psychologie ordinaire, ne peut qu'être intensionnelle et, par là même, elle n'instancie aucune loi déterministe stricte au sens de la physique. D'où son caractère non nomologique ou « anomal ». En bref, il y a donc une autonomie explicative des concepts psychologiques ou mentaux, mais, ontologiquement, la causalité comme relation ne fait pas intervenir le mental en tant que mental. Celui-ci reste un épiphénomène – on n'ose dire une « superstructure ».

Rapprochant diverses tentatives de donner corps à un matérialisme naturaliste, dont le fonctionnalisme et le monisme anomal, Pascal Engel écrit : « Ce double souci de réduction des concepts mentaux à des concepts acceptables du point de vue d'une psychologie scientifique et de maintien d'une autonomie de ces concepts illustre en fait le dilemme permanent d'une théorie matérialiste de l'esprit. En effet, plus la réduction est réussie, c'est-à-dire plus on parvient à « expliquer » les concepts mentaux en termes « physicalistes » ou « naturalistes », moins nos concepts mentaux usuels, ceux de la psychologie du sens commun et de notre conception préthéorique de l'esprit apparaissent corrects, et plus on est tenté d'« éliminer » les seconds au profit des premiers, et de considérer qu'il n'y a tout simplement pas de croyances, de désirs, de sensations, etc. En d'autres termes, le matérialisme oscille sans cesse entre ses versions « éliminativistes » et ses versions « non réductionnistes ». Le projet d'une philosophie de l'esprit « naturalisée » oscille pareillement entre ces deux tendances[13] ». A quoi on peut ajouter que les versions non réductionnistes flirtant bien souvent avec des formes déguisées de dualisme, l'instabilité soulignée par Engel est encore plus forte qu'il ne le dit. C'est ce que souligne John Searle à propos de l'intelligence artificielle et du fonctionnalisme : « L'intelligence artificielle dans sa version forte ne prend sens que par rapport à l'hypothèse dualiste selon laquelle, pour qui s'intéresse à l'esprit, le cerveau n'a pas d'importance. En intelligence artificielle, aussi bien que pour le fonctionnalisme, ce qui compte, ce sont les programmes, et les programmes sont indépendants de leur réalisation dans des machines [...]. C'est seulement si l'on croit que l'esprit est séparable du cerveau à la fois conceptuellement et empiriquement – c'est-à-dire si l'on est dualiste dans un sens fort – que l'on peut former l'espoir de reproduire le mental en écrivant et en faisant tourner des programmes, puisque les programmes doivent être indépendants des cerveaux ou de toute autre forme d'instantiation[14] ». Quine, qui, pour sa part, en dépit d'une philosophie personnelle proche du matérialisme éliminationniste, soutient le monisme

anomal de Davidson, souligne tout ce que ce « monisme » comporte de dualisme : « L'ancien dualisme de l'esprit et du corps peut donc subsister, transmué et transplanté en un dualisme des concepts ou du langage. Comme tel, il demeure irréductible[15]. »

Ironie de l'histoire, une fois encore. Rétrospectivement, on peut dire que la naturalisation de la philosophie transcendantale était, paradoxalement, peut-être plus facile à mener à bien que la naturalisation d'une philosophie psycho-linguistique de l'esprit ; et que cette tâche était à la portée de la cybernétique, mais qu'elle n'a pas su s'en donner les moyens.

Brentano trahi

Pour illustrer cette double assertion, il est nécessaire de faire une brève incursion dans l'autre grand courant de la philosophie du xxe siècle, à savoir la phénoménologie. On sait, bien sûr, que le concept d'intentionnalité y joue un rôle décisif, et on se doute qu'il n'a, au mieux, que des rapports très lointains avec ce que le même mot désigne en philosophie cognitive, puisque, chez Husserl, ce sont les « vécus psychiques » qui ont la propriété d'être « intentionnels », cette propriété constituant l'« essence du concept de conscience ». Ce qu'on sait généralement moins, c'est que l'intentionnalité en philosophie de l'esprit et l'intentionnalité husserlienne ont un ancêtre commun, à savoir la psychologie philosophique de Franz Brentano. Husserl fut l'élève de celui-ci, à Vienne, de 1884 à 1886, et l'influence qu'il en reçut fut décisive, comme en attestent les *Recherches logiques* de 1901. Quant à Roderick Chisholm, avant d'être, comme on l'a dit, l'inventeur de la version linguistique de l'intentionnalité, il fut l'introducteur de la pensée de Brentano sur le sol américain, traducteur et commentateur de plusieurs de ses livres, bref, l'une des autorités brentaniennes les plus incontestées. Et pourtant, de Husserl et de Chisholm, il faut bien que l'un des deux au moins ait trahi son maître.

« Toute conscience est conscience de quelque chose » : cette formule de Husserl, dont on sait, en France, ce qu'elle est devenue dans cette variante de la phénoménologie qu'est l'existentialisme sartrien, c'est de la pensée de Brentano que Husserl la tire, et plus précisément de ce passage, qu'il cite, de l'ouvrage du maître de Vienne, publié en 1874 sous le titre : *Psychologie du point de vue empirique*[16] : « Tous les phénomènes psychiques sont caractérisés par ce que les scolastiques du Moyen Age ont appelé l'*inexistence intentionnelle* (ou encore mentale) d'un objet, et que nous pourrions appeler, quoique de façon non dénuée d'ambiguïté, la relation à un contenu, l'orientation vers un objet (terme qui ne doit pas être compris ici comme signifiant une chose), ou l'objectivité immanente[17]. »

Tous les mots, ici, sont des pièges en puissance. « Intentionnel », nous rappelle Brentano, est un terme scolastique qu'utilise, par exemple, Thomas d'Aquin dans le sens de « mental », par opposition à « réel ». Dans l'objet réel, c'est-à-dire situé en dehors de l'esprit, la forme est unie à la matière ; l'objet « intentionnel », lui, n'est présent que par sa forme. *Inexistenz* (« inexistence ») vient du latin *in-esse*, qui signifie : « être à l'intérieur de ». Le contresens serait ici, évidemment, de comprendre « inexistant » comme voulant dire « non existant ». L'objet *vers* lequel *tend* l'esprit (son *intention*) se situe *à l'intérieur* de l'esprit ; voilà pourquoi sa présence est « immanente ».

La vie psychique, selon Brentano, est avant tout une activité, un processus, une

dynamique. Cette activité est « présentation ». Brentano précise : « Par présentation, je ne veux pas dire cela même qui est présenté, mais bien plutôt l'activité de présentation. » Cette activité a un contenu, ou plus précisément un « objet ». L'objet est cela même qui est présenté : le son que nous entendons, la couleur que nous voyons, le froid que nous ressentons. Ces objets de la présentation, Brentano les nomme, comme pour mieux brouiller les pistes, « phénomènes physiques » – par opposition à l'activité psychique précédemment définie. Mais il précise, pour enlever toute ambiguïté, que ces phénomènes « physiques » font bien partie des « données de la conscience ».

L'activité psychique est intrinsèquement consciente d'elle-même. Quand nous pensons, nous avons une perception immédiate du fait que nous pensons, et la perception de l'activité pensante est simultanément perception de l'objet de la pensée. Cette perception interne ne peut pas être une observation, note Brentano, car il y aurait alors régression infinie d'activités psychiques pointant les unes vers les autres. C'est en une appréhension globale et unique que la pensée comme activité se rapporte à la fois à elle-même et à son objet intentionnel : « *Percevons*-nous les phénomènes psychiques qui existent à l'intérieur de nous ? On doit répondre emphatiquement oui à cette question ; car d'où nous viendraient les concepts de présentation et de pensée sans une telle perception ? En revanche, il est évident que nous ne pouvons pas *observer* les phénomènes psychiques en nous [...]. Cela suggère qu'il y a un lien particulier entre l'objet de la présentation interne et la présentation elle-même, et que l'un et l'autre relèvent d'une seule et même activité psychique. La présentation d'un son et la présentation de la présentation d'un son relèvent d'un seul et même phénomène psychique ; c'est seulement lorsque nous considérons ce dernier dans sa relation à deux objets différents, l'un d'entre eux étant un phénomène physique et l'autre un phénomène psychique, que nous le divisons conceptuellement en deux présentations [...] C'est dans le phénomène psychique même par lequel le son se présente à l'esprit que nous appréhendons simultanément le phénomène psychique lui-même[18]. »

On sait ce que Husserl fera de cette idée. Il en déduira le geste phénoménologique par excellence, l'*epoché* ou « réduction », qui revient à décider de ne voir dans l'objet de la représentation que le corrélat (dit « noématique ») de l'activité psychique (dite « noétique »), mettant entre parenthèses le problème du rapport entre l'objet de la représentation (l'objet intentionnel) et la chose en soi. L'intentionnalité ainsi comprise est, ainsi que le montre excellemment Alain Renaut, l'arme principale de Husserl dans sa critique du psychologisme : « Dire que, pour qu'il y ait conscience, il faut que l'objet et le sujet se rapportent l'un à l'autre d'une manière telle que l'objet vienne à « in-exister intentionellement » pour le sujet (c'est-à-dire à prendre la forme de l'objet de la représentation, distinct de ce qu'il est hors de la conscience), c'était en effet aussi désigner une relation entre objet et sujet qui n'est plus l'affaire de la psychologie [...]. Il existe, à l'égard de nos représentations, un niveau de questionnement plus radical que celui de la psychologie : là où cette dernière présuppose toujours-déjà le fait que nous avons des représentations, pour se borner à montrer le rôle qu'y jouent la perception, la mémoire, l'attention ou les sentiments, il s'agit de poser la question pré-psychologique du *fait* même qu'il y a des représentations et du *mode d'être* de l'objet de la représentation – question *ontologique*, si l'on veut, au sens où il y va en elle de l'existence et de l'essence des représentations, question *transcendantale*, si l'on préfère, au sens où il y va en elle des conditions de possibilité de la représentation[19]. »

Imaginons maintenant un (mauvais) élève de philosophie de première année lisant les textes de Brentano que nous avons cités et se méprenant systématiquement sur le sens des termes clés. L'objet vers lequel tend la représentation est un *objet physique inexistant.* C'est donc la chose en soi. Quand je pense à une vache ou à un dahu, qu'est-ce qui, en effet, est absent de ma représentation ? Le dahu en chair et en os, si l'on peut dire, parce qu'il n'existe pas, et la vache avec ses cornes et son lait, parce qu'elle existe, mais en dehors de mon esprit. L'intentionnalité, ce n'est donc plus l'*activité* psychique se dépassant elle-même à l'intérieur d'elle-même en direction d'un objet qui lui reste intérieur, cette « transcendance dans l'immanence » que tente de cerner Husserl ; cela devient un *état* mental doté d'un contenu, lequel se rapporte à un objet dont l'existence n'est pas garantie par le fait que l'état mental, lui, existe. Le contenu ne peut être qu'intensionnel, donc linguistique.

Si énorme que cela puisse paraître, c'est dans ce piège interprétatif que, « consciemment » ou « inconsciemment », Roderick Chisholm est tombé, ou, en tout cas, a fait tomber ses lecteurs philosophes analytiques, le premier et le plus célèbre d'entre eux étant Willard Van Orman Quine. Le dossier est maintenant bien établi[20] . Dans une section, qui devait exercer une forte influence sur ses nombreux et, à leur tour, influents lecteurs, de son livre de 1960, *Word and Object*[21] , Quine qui, de toute évidence, n'a pas lu la *Psychologie vom empirischen Standpunkt*, s'en remet à l'« éclairante » lecture de Chisholm pour attribuer à Brentano une « thèse » avec laquelle il se dit en accord – puisqu'elle ne fait qu'un, selon lui, avec sa propre thèse de l'indétermination radicale de la traduction d'une langue dans une autre, ou des contenus mentaux qu'un auditeur attribue à un locuteur. Cette « thèse de Brentano » est, bien sûr, la thèse de Chisholm. Elle affirme que les états mentaux, et eux seuls, sont dotés de la propriété d'intentionnalité, entendue comme rapport de type linguistique à des objets ou états de chose du monde extérieur à l'esprit. Les expressions intentionnelles sont irréductibles aux termes que nous utilisons pour décrire les phénomènes « physiques ». « On peut accepter la thèse de Brentano soit comme montrant le caractère indispensable des expressions intentionnelles et l'importance d'une science autonome de l'intention, conclut Quine, soit comme montrant le caractère non fondé des expressions intentionnelles et la vacuité d'une science de l'intention. Mon attitude, contraire à celle de Brentano, est la seconde[22]. » Le tour était donc joué. D'autres, moins « éliminationnistes » que Quine, plus conscients des limites de toute entreprise de « naturalisation de l'épistémologie », allaient concevoir des « monismes non réductionnistes » capables de concilier la « thèse de Brentano » avec un physicalisme relatif. On connaît la suite.

On aurait tort de se gausser. L'histoire de la philosophie illustre assez bien l'assertion que les philosophes sont, les uns pour les autres, leurs plus mauvais lecteurs ; mais aussi que les bévues et les contresens qu'ils commettent sont parfois l'occasion d'œuvres riches, originales et faisant école. C'est la seule chose qui doit retenir l'historien des idées. Pour en rester à l'histoire de la phénoménologie, lorsque Sartre attribue à Heidegger la thèse « qu'il faut partir de la subjectivité[23] », la méprise n'est pas moins cruelle que celles commises par Chisholm et par Quine…

La rencontre ratée avec la phénoménologie

Daniel Andler a raison de souligner qu'on trouverait difficilement dans la cybernétique une notion de représentation qui préfigurerait tant soit peu la théorie psycho-linguistique qui constitue aujourd'hui le noyau dur de la philosophie cognitive. Est-ce à dire que la question transcendantale « du fait même qu'il y a des représentations et du mode d'être de l'objet de la représentation » échappait aussi au projet cybernétique ?

Avec le recul des années, on peut dire que s'il existe une possibilité de modéliser scientifiquement – et donc, si l'on veut, de « naturaliser » – l'intentionnalité telle que l'a conçue Brentano et, à sa suite, Husserl, c'est bien l'étude des réseaux telle que l'a lancée Warren McCulloch qui l'a ouverte. Cependant, pour que la cybernétique en formât le projet, il eût fallu qu'elle prît sur le réseau de neurones idéalisés un tout autre point de vue que le sien.

McCulloch et Pitts, on y a insisté, voyaient leurs réseaux comme des machines logiques. Pénétrés du caractère révolutionnaire de la thèse de Turing, forts du théorème d'équivalence entre leur machine et celle de Turing, la mission éducative qu'ils s'étaient assignée était de montrer que, pour chacune des grandes facultés de l'esprit, on pouvait concevoir un réseau capable de la reproduire. Or la conception turingienne d'une fonction de l'esprit en fait une fonction au sens mathématique, c'est-à-dire un opérateur transformant des entrées en des sorties (des stimuli en des réponses). Cette perspective, en définitive behavioriste, constituait un obstacle au *Gestalt switch* qui eût permit de traiter le réseau comme un système dynamique « autonome », c'est-à-dire informationnellement et organisationnellement clos, sans entrée ni sortie. Le changement de perspective ne devait se produire que beaucoup plus tard, au confluent de divers courants ou écoles de pensée : issues de la seconde cybernétique, les tentatives de formaliser l'auto-organisation des systèmes biologiques au moyen de réseaux d'automates booléens (en France, ces travaux furent principalement menés, dans les annéees soixante-dix et quatre-vingt, par une équipe dirigée par Henri Atlan, comprenant Françoise Fogelman-Soulié, Gérard Weisbuch et Maurice Milgram[24] ; aux États-Unis, à l'institut de Santa Fe pour l'étude des systèmes complexes, par Stuart Kauffman[25] , ancien élève de Warren McCulloch) ; également issue de la seconde cybernétique, l'école chilienne de l'autopoïèse, qui va si loin dans sa conception de la clôture informationnelle qu'elle en nie l'existence même des représentations (Humberto Maturana et Francisco Varela sont les deux fondateurs de ce courant qui reçut la bénédiction de Gregory Bateson, Varela[26] recourant à divers formalismes relevant de la théorie des réseaux d'automates) ; au cœur même du néoconnexionnisme, se distinguant du courant « PDP » déjà signalé (lequel prolonge, en les complexifiant, les recherches de type « Perceptron »), le courant « ANN » (pour *Attractor Neural Network*), animé par des physiciens, qui étudie les propriétés émergentes de réseaux (presque) complètement connectés, l'information se propageant donc dans toutes les directions (fondé par John Hopfield[27] , ce courant connaît des développements remarquables dans les travaux de Daniel Amit[28] en Israël ; en France, les recherches de Jean Petitot[29] appliquant la théorie des systèmes dynamiques à l'étude des réseaux d'automates se rattachent également à ce courant).

Tous ces travaux, dans leur foisonnement, ont un point commun. Ils traitent un réseau complexe de calculateurs élémentaires en interaction comme un être « autonome », on l'a dit, au sens où, doté d'une spontanéité propre, il est à lui-même la source de ses déterminations, et non le simple transducteur convertissant des messages d'entrée en messages de sortie. Du modèle de McCulloch et Pitts à celui de Daniel Amit, l'être

mathématique n'a pas été bouleversé : il reste, fondamentalement, l'automate à seuil. C'est la perspective que l'on prend sur lui qui change du tout au tout. On s'intéresse non plus à ses capacités computationnelles, mais à ses « comportements propres » (*eigenbehaviors*, dans l'anglo-germanique de la mécanique quantique à qui la théorie des systèmes a emprunté ce concept). Qu'est-ce à dire ? Comme tout automate à état interne, un réseau calcule son état à l'époque suivante en fonction de son état à l'époque présente. Or une propriété très générale qui caractérise les réseaux est que, après une période de transition souvent assez courte, le comportement collectif se stabilise en un « cycle limite » (c'est-à-dire une configuration spatio-temporelle périodique) de faible périodicité (la période pouvant être égale à un, auquel cas on a affaire à un état stationnaire, ou point fixe). Tout se passe comme si ce comportement collectif stable était *auto*-reproducteur, c'est-à-dire se produisait lui-même – d'où l'expression « comportement propre », *self-behavior* –, alors qu'en vérité il reste produit par le réseau. Considérons, par exemple, le cas d'un automate élémentaire, ou neurone, qui n'est jamais mis à feu au cours du cycle limite. Il semblerait ainsi qu'il n'ait aucun effet causal sur la détermination de celui-ci. Ce n'est qu'une illusion, et il suffirait de le retirer, lui et ses synapses, de la structure du réseau pour qu'on s'aperçoive que le comportement collectif en serait affecté.

Un réseau donné possède en général une multiplicité de comportements propres – on dit aussi « attracteurs », terme emprunté à la théorie des systèmes dynamiques –, et il convergera vers l'un ou l'autre d'entre eux en fonction de ses conditions initiales. La « vie » d'un réseau peut ainsi se concevoir comme une trajectoire dans son « paysage » d'attracteurs, le passage de l'un à l'autre résultant de perturbations ou chocs en provenance du monde extérieur. On convient de dire que ce sont là des événements significatifs pour le réseau, et que le *contenu* de sens qu'il leur attribue est précisément le comportement propre ou attracteur qui en résulte. C'est évidemment un contenu purement endogène, et non pas le reflet d'une objectivité extérieure, « transcendante ».

On voit où nous voulons en venir et où certains, déjà, ont planté leur chevalet. N'a-t-on pas ici, au moins en germe, un très beau modèle de l'« objectivité immanente » dont parlait Brentano ? L'attracteur est simultanément un être qui participe pleinement de l'activité du réseau et qui, cependant, en un sens la transcende, puisqu'il relève d'un niveau supérieur de complexité logique. La dynamique du réseau *tend vers* un attracteur, mais celui-ci n'est qu'un produit de la dynamique du réseau. Le réseau est un être *intentionnel*, au sens de Brentano et de Husserl. La théorie des systèmes inventera d'ailleurs un terme pour parler de ce rapport paradoxal entre une dynamique et son attracteur. Elle parlera d'« autotranscendance ». La « transcendance dans l'immanence » de Husserl n'est pas loin. On ne s'étonne pas que certains chercheurs en sciences cognitives, considérés, évidemment, comme « marginaux », et qui utilisent dans leurs travaux les réseaux d'automates, se recommandent, de près ou de loin, d'une phénoménologie transcendantale. (En France, on peut songer aux recherches d'Henri Atlan[30] , de Jean Petitot[31] et de Francisco Varela[32].)

Dès l'époque des conférences Macy, la cybernétique eût-elle pu emprunter cette voie ? J'ai cité l'obstacle idéologique que constituait le paradigme turingien. Il y a aussi les circonstances, en particulier la formation et l'information philosophiques des protagonistes.

Le seul philosophe professionnel à participer aux conférences Macy était Filmer Northrop, professeur à Yale et jouissant à l'époque d'une certaine réputation. Dans les

années trente, il avait animé un séminaire à l'intention des scientifiques intéressés par la philosophie, où s'étaient rencontrés plusieurs des futurs participants aux conférences, dont McCulloch. Mais la philosophie personnelle de Northrop était scientiste. Il cherchait à réconcilier les divers systèmes philosophiques du monde, passé et présent, grâce à la science, et plus spécialement, lorsqu'il les découvrit, grâce aux modèles de Rosenblueth-Wiener-Bigelow et de Pitts-McCulloch. Son influence sur le groupe fut des plus réduites. La formation philosophique des cybernéticiens était essentiellement réduite à la logique philosophique. Wiener avait été l'élève de Russell à Cambridge, von Neumann disciple de Hilbert à Göttingen, Pitts avait étudié la logique symbolique avec Carnap, à Chicago, en 1938. McCulloch était grand lecteur de Russell, de G.E. Moore, de Peirce, et du Wittgenstein du *Tractatus*[33].

La rencontre avec la phénoménologie husserlienne, ou plutôt l'un de ses dérivés, eut lieu, cependant, mais ce fut sur le mode d'une confrontation, d'ailleurs ratée, avec la psychologie de la *Gestalt*. Celle-ci, une création de l'Allemagne de Weimar, avait été introduite aux États-Unis par la vague d'immigration accompagnant la montée du nazisme. Wolfgang Köhler avait été l'un des fondateurs du mouvement, dès les années vingt, dans le cadre de l'institut psychologique de Berlin, dont il avait repris la direction à son maître, Carl Stumpf, lui-même ancien élève de Brentano et proche de Husserl. Le projet de l'équipe, qui comprenait aussi Kurt Koffka, Max Wertheimer et Kurt Lewin, était de résoudre les problèmes que la phénoménologie husserlienne avait posés au moyen d'une psychologie scientifique de type expérimental, en s'appuyant sur les concepts de la physique quantique, et en particulier celui de champ. L'ambition était de rechercher les lois, au sens que ce mot a dans les sciences de la nature, qui gouvernent la perception et l'expérience immédiate que nous avons des choses, en s'attachant avant tout à sauver leur caractère « holiste » de totalités organisées.

Dès les années vingt, Köhler avait postulé un « isomorphisme psycho-physique » entre les événements physiques survenant dans le cerveau et les faits psychologiques. L'isomorphisme qu'il cherchait à cerner était de type topologique : la perception interne d'une figure sur son fond, par exemple, devait avoir, pensait-il, son correspondant géométrique dans le cerveau. Une fois aux États-Unis, Köhler allait procéder à des expériences neurophysiologiques dans l'espoir de découvrir un corrélat physique aux phénomènes de perception sous la forme de champs électriques continus dans le tissu cérébral. Lorsque les cybernéticiens l'invitèrent à participer à la quatrième conférence Macy, en 1947, ils s'attendaient à ce que la discussion portât-comme ils en avaient déjà fait l'expérience avec d'autres représentants d'une conception continuiste du fonctionnement cérébral – sur les mérites comparés de cette dernière et du modèle de neurones idéalisés. En fait, la discussion tourna court, Köhler se cantonnant dans un discours qui fut jugé purement idéologique[34]. Cette occasion manquée fut partiellement rattrapée, l'année suivante, au symposium Hixon, où la confrontation entre le mécanisme de McCulloch et le holisme de Köhler prit néanmoins des allures de guerre de religions.

Entre le husserlien Wolfgang Köhler et l'atomiste logique Warren McCulloch, celui qui se rapprocha en définitive le plus d'une naturalisation de l'enquête transcendantale ne fut paradoxalement pas le premier. Köhler croyait au parallélisme géométrique entre la perception et son substrat matériel ; McCulloch enquêtait, lui, sur les conditions de possibilité formelles et matérielles de toute connaissance. L'un de ses plus beaux succès en la matière fut le travail qu'il réalisa sur la grenouille, bien plus tard, avec Walter Pitts,

Jerome Lettvin et Humberto Maturana (c'est de ce travail qu'allait sortir l'école chilienne des « systèmes autopoiétiques », cet autre fleuron de la seconde cybernétique). Cette recherche expérimentale, publiée sous le titre éloquent : « What the Frog's Eye Tells the Frog's Brain » (« Ce que l'œil de la grenouille dit au cerveau de la grenouille »)[35], mettait en évidence l'existence de récepteurs dans le système visuel répondant sélectivement à certains traits du stimulus, comme la convexité, et les interprétait comme les équivalents matériels des catégories du jugement synthétique *a priori*. Peu de temps après, Hubel et Wiesel devaient confirmer ces résultats sur le chat, ce qui leur valut le prix Nobel de médecine.

Une philosophie de l'esprit sans sujet

Kantisme, oui, mais sans sujet transcendantal, disait Paul Ricœur à propos du structuralisme de Claude Lévi-Strauss[36] . La formule s'applique à merveille à la cybernétique. Pas plus que le « Système symbolique physique » de Newell et Simon, le réseau neuronal de McCulloch et Pitts, n'est, ni ne prétend être, un modèle du sujet. De l'esprit (au sens de *mind*), oui, du sujet, non. Un esprit sans sujet, voilà peut-être l'apport le plus significatif de la cybernétique à la philosophie. Le rapprochement avec Lévi-Strauss s'impose d'autant plus que celui-ci faisait explicitement référence à la cybernétique dans son « Introduction à l'œuvre de Marcel Mauss » que beaucoup considèrent comme le premier manifeste du structuralisme français. La tâche de l'anthropologie sociale, affirmait-il, est de révéler les « structures mentales inconscientes » à l'œuvre derrière des pratiques comme celle de l'échange de dons dans les sociétés archaïques. Ces structures, qui jouent le rôle d'un synthétique *a priori*, se manifestent dans des phénomènes de communication, et la « pensée symbolique » est structurée comme un langage[37]. Nous sommes en 1950, et Lévi-Strauss a déjà opéré son tournant sémiotique sous l'influence de Roman Jakobson. Or, les deux années qui précèdent, ont paru coup sur coup *Cybernetics* de Norbert Wiener (1948) et *The Mathematical Theory of Communication* de Claude Shannon et Warren Weaver (1949). Lévi-Strauss se persuade qu'une théorie mathématique de la communication apportée par la linguistique à l'anthropologie est la voie d'avenir pour les sciences sociales…

En inventant un type d'enquête transcendantale qui faisait l'économie du sujet, la cybernétique aura bien servi la déconstruction de la métaphysique de la subjectivité. C'est là un point essentiel qui est trop rarement vu. Dans une interview restée célèbre, parue en 1976 dans le numéro 23 de *Der Spiegel*, Heidegger lançait l'anathème : « La cybernétique est la métaphysique de l'âge atomique. » Pour la plupart des philosophes de la mouvance heideggérienne, mais aussi pour beaucoup d'autres, la cybernétique apparaît comme le comble de la « philosophie du *cogito* », l'aboutissement d'une métaphysique qui confie à l'homme le projet de se rendre maître et possesseur de toutes choses. La cybernétique, summum de la volonté de puissance et de la volonté de volonté, nous ferait passer « à côté de l'essence de la technique » – qui est, comme on sait, à chercher hors de la technique, dans ce mode de révélation et de dévoilement de l'Être que constitue l'« arraisonnement ». « Aussi longtemps que nous nous représentons la technique comme un instrument, constatait Heidegger dans sa fameuse conférence sur l'essence de la technique prononcée pour la première fois en 1949 devant le Club de Brème[38], nous

restons pris dans la volonté de la maîtriser. » Cette définition instrumentale et anthropologique de la technique serait donc portée à son comble par l'avènement de la cybernétique.

Ceux qui vont répétant cette condamnation ne se sont, semble-t-il, jamais interrogés sur ceci, qui aurait pourtant dû les troubler. Lorsque, à partir du début des années cinquante, les « sciences de l'homme » à la française (structuralistes puis post-structuralistes), dont la filiation avec la pensée de Heidegger est indiscutable, ont eu à dire le mystèrende l'être et son dévoilement, le langage poétique n'étant pas nécessairement leur fort, c'est dans la cybernétique qu'elles sont allées puiser leurs métaphores. S'agissait-il de montrer que l'homme n'est pas maître chez lui, que loin de maîtriser son langage, c'est celui-ci qui le maîtrise, et que l'« ordre symbolique » est irréductible à l'expérience humaine – bref, que « ça parle » lorsque l'homme croit parler ? C'est aux automates cybernétiques et à la théorie de l'information que l'on a fait systématiquement référence. Le cas le plus typique est celui de Jacques Lacan qui plaça son séminaire entier de l'année 1954-1955 sous le signe de la cybernétique, l'apothéose en étant la lecture de *The Purloined Letter* d'Edgar Poe dans les termes de la théorie des automates[39] . Lacan devait plus tard ouvrir ses *Écrits* par une réélaboration de ce séminaire sur *La Lettre volée,* texte obscur dont il fournissait, prétendait-il, la clé dans une « Introduction » – paradoxalement rejetée à la fin –, qui prenait la forme de variations cybernétiques sur le jeu de pair ou impair destinées à illustrer « comment un langage formel détermine le sujet ». Même si cette partie de l'enseignement du maître semble être passée au-dessus de la tête de plus d'un de ses disciples, on ne saurait en conclure que Lacan ne prenait pas la cybernétique très au sérieux. En témoigne également la conférence « Psychanalyse et cybernétique, ou de la nature du langage », prononcée en juin 1955. Plus tard, la nouvelle critique littéraire, grisée à son tour par l'autonomie des signifiants et leur dérive infinie, allait actionner avec dextérité divers types de « machines textuelles », désirantes ou non. Il n'est pas jusqu'à la « mise en abyme » et autres formes d'autoréférence littéraire qui ne devaient chercher une formalisation dans les concepts de la cybernétique[40].

Il est de bon ton aujourd'hui d'affirmer qu'il n'y avait là qu'effet superficiel de mode. Dans le cas de Lacan, c'est assurément faux, et celui-ci avait, sur certains points, de la cybernétique une connaissance fort précise. Il s'intéressait, par exemple, comme nous l'avons dit, à la théorie des circuits fermés réverbérants que McCulloch avait reprise à Lawrence Kubie, et connaissait les travaux du neuro-anatomiste britannique John Z. Young visant à tester cette théorie sur le poulpe (ces travaux furent discutés lors de la neuvième conférence Macy, en 1952)[41] . Les recherches de la seconde cybernétique ne lui étaient pas non plus étrangères, puisqu'il avait compris ce qu'il pouvait retirer des travaux formels de Ross Ashby, dont nous reparlerons, sur les propriétés différenciatrices du calcul[42].

On retrouve, en réalité, dans le projet cybernétique cette ambiguïté dont parle Heidegger à propos de l'essence de la technique[43] . La technique révèle la vérité (heideggérienne) au sujet de l'Être, vérité qui inclut la déconstruction de la conception métaphysique du sujet ; mais, simultanément, son mode particulier de dévoilement - l'« arraisonnement » – fait courir à l'humanité et au monde un danger, celui de l'engloutissement dans un projet frénétique de puissance et de maîtrise. C'est une même ambiguïté que note Philippe Breton, l'un des très rares historiens du destin de la cybernétique en France : « La cybernétique a été l'un des principaux instruments

déstabilisateurs de la conception anthropocentrique de l'homme [...]. La cybernétique assume donc un terrible paradoxe, celui d'affirmer l'humanité tout en en dépossédant l'homme. En ce sens, elle exprime peut-être à découvert un trait fondamental de la connaissance scientifique et technique contemporaine, qui fait que les bénéfices du progrès semblent irrémédiablement associés avec la mise en scène rationnelle de la mort de l'homme[44] . » Jean-Claude Beaune trouve également des accents heideggériens lorsque, dans son livre *L'Automate et ses mobiles*[45] , il écrit : « Ambiguïté qui rappelle l'essence mythique de la technicité automatique : plus l'imitation technique du modèle (l'homme) est fidèle, plus l'anthropocentrisme règne, plus les finalités humaines de l'objet ont besoin d'être nettement et brutalement énoncées, car l'objet tend de plus en plus à échapper à son fabricant. Plus l'homme met dans sa production l'image humaine, plus cette image lui échappe. » Avec l'automate cybernétique, « une nouvelle révolution copernicienne s'amorce : le centre de gravité du monde n'est plus l'homme, mais la machine ». Pour sa part, Gilbert Hottois, reprenant le thème de l'« automatisation » de la technique développé par Jacques Ellul, croit y voir un accès à la découverte de « l'essence non anthropologique de la technique », de sa « nature fondamentalement *ab-humaine*[46] ».

C'est sans doute, en effet, dans le sens donné aux recherches sur les automates que l'ambiguïté du projet cybernétique apparaît le mieux. On peut voir en elles le comble de l'anthropomorphisation de la technique. C'est ce à quoi le public est le plus sensible ; mais ce fut aussi une des critiques favorites de Heinz von Foerster à l'égard de la première cybernétique (et de l'intelligence artificielle qui prit sa suite) que de l'accuser d'avoir parlé des machines en termes anthropomorphiques. On peut cependant tout aussi bien choisir de voir dans l'automate cybernétique la mécanisation de l'humain, le dévoilement de ce qu'il y a de non humain en l'homme. Ainsi, Lacan : « La question de savoir si [la machine] est humaine ou pas est évidemment toute tranchée – elle ne l'est pas. Seulement, il s'agit aussi de savoir si l'humain, dans le sens où vous l'entendez, est si humain que ça[47]. »

Dans son étude *Le Cerveau-machine*, au titre (volontairement ?) bien mal choisi, puisqu'il contredit la philosophie « humaniste » de l'ouvrage, Marc Jeannerod caractérise parfaitement l'esprit de la cybernétique, même si c'est négativement et sur un ton réprobateur, lorsqu'il écrit, évoquant l'existence maintenant démontrée de cellules nerveuses qui, telles des horloges biochimiques, possèdent une activité rythmique spontanée : ces générateurs d'oscillations spontanées « permettent de ramener le fonctionnement du système nerveux à celui d'une machine qui « marche toute seule », une fois déclenchée, ou qui pourrait même, à la limite, se déclencher elle-même. Une telle physiologie de la spontanéité reste cependant une explication singulièrement limitée du comportement. Si elle rend bien compte de ses aspects automatiques, elle ne pourrait par contre être généralisée à ses autres aspects sans aboutir à un *comportement sans sujet*[48] ». Mais tel était précisément l'objectif des cybernéticiens. Ils participaient bien à leur manière à la critique de la métaphysique de la subjectivité.

C'est parce qu'ils étaient, à juste titre, sensibles à cet aspect que les heideggériens français ont, de façon cohérente, fait appel à la cybernétique pour exprimer ce qu'ils avaient à dire. Lorsque, dans son étude par ailleurs excellente de la philosophie française de l'après-guerre, Vincent Descombes reproche au structuralisme de faire preuve d'incohérence en prétendant lutter contre « la philosophie de la conscience » tout en puisant ses concepts dans « une pensée d'ingénieurs », il ne fait que manifester une

méconnaissance de la cybernétique qui est largement partagée – et que le choix, sans doute malheureux, du mot même de cybernétique (théorie de la commande, de la gouverne, de la maîtrise) ne saurait excuser[49].

McCulloch *vs* Wiener

Je voudrais conclure cette enquête sur la dimension philosophique de la cybernétique par une brève comparaison entre les styles de pensée de Norbert Wiener et de Warren McCulloch.

Le premier a donné son nom à la cybernétique, et il a façonné son image de marque par des livres qui eurent la faveur du grand public : *Cybernetics* (1948) et surtout *The Human Use of Human Seings* (1950). Il a aussi fixé son vocabulaire de base (information, message, communication, « contrôle », feedback, etc.) et contribué à lui donner un certain style, que nous dirions aujourd'hui « écologique » et qui manifeste la volonté de respecter l'interdépendance des phénomènes. En regard, McCulloch paraît une figure effacée, et nombreux aujourd'hui sont ses héritiers qui ignorent jusqu'à son nom. Toute histoire sérieuse de la cybernétique se doit d'abord de rétablir l'équilibre entre les deux hommes, surtout si elle se fonde, comme c'est le cas ici, sur l'étude des conférences Macy, tout au long conçues, organisées et animées par McCulloch.

Autant Wiener était brouillon dans son travail de penseur, abusant des métaphores et des analogies, autant McCulloch fut rigoureux dans la poursuite obstinée de sa quête philosophique. Le contraste entre les deux hommes apparaît nettement dans leur rapport à la logique. Même s'il reconnaît la grande influence qu'eut sur lui Russell[50] , Wiener n'a jamais été un inconditionnel de la logique formelle. Très tôt, ce que Steve Heims nomme chez lui un certain « romantisme typique du XIXe siècle » le convainquit que toute tentative d'enserrer le monde des phénomènes dans un système de connaissance clos sur lui-même et parfaitement logique est vouée à l'échec. Notre connaissance du monde doit laisser leur place au paradoxe, à l'incomplétude, au hasard. La publication du théorème de Gödel en 1931 ne fut donc pas pour lui une révélation[51]. L'itinéraire de McCulloch devait en revanche, comme on l'a vu, le conduire à s'enfermer de plus en plus dans l'univers de la logique.

L'un et l'autre sont apparus comme des marginaux aux yeux des communautés scientifiques de l'époque, mais pour des raisons fort différentes. C'est l'éclectisme de Wiener qui déconcertait ses collègues : muni de sa boîte à outils cybernétiques, il pouvait démonter n'importe quel mécanisme descriptible en termes de comportement ou de communication[52] . De là, sans doute, l'image d'« ingénieur » qui s'attache à lui, mais qui n'est qu'une mauvaise approximation de son style. McCulloch fut par contraste l'homme d'une idée fixe – tenir ensemble, quoi qu'il en coûte, le cerveau, l'esprit et la machine –, mais cette idée était trop « philosophique » pour que les physiologistes de son temps pussent la comprendre[53]. Il ne se considérait cependant pas lui-même comme philosophe. Il voulait édifier une science naturelle (logique) de l'esprit, et c'est dans le cadre de la méthode scientifique qu'il cherchait une réponse aux vieilles questions philosophiques qui le tourmentaient. C'est cette science qu'il avait baptisée du nom d'« épistémologie expérimentale ».

La rigueur ou l'éclectisme ne sont certes pas des facteurs décisifs dans l'influence

qu'exerce un penseur. S'il est cependant possible de soutenir que McCulloch, mieux que Wiener, incarne, tant dans sa personne que dans son œuvre, ce qui se joue dans le projet cybernétique, c'est qu'on trouve chez le premier, mais non chez le second, une profonde cohérence entre le message idéologique et l'œuvre scientifique. L'un comme l'autre auront, en fin de compte, bien servi la « déconstruction » de la conception leibnizienne et cartésienne du sujet ; Wiener en soutenant que la volonté est de l'ordre du mécanisme, McCulloch en faisant de même avec la perception, la pensée et la conscience. Grâce à eux, il est désormais possible de donner des représentations rigoureuses de la notion de *processus* (comportement ou pensée) *sans sujet*. Or, même s'il ne recourait pas à cette terminologie heideggérienne, McCulloch, lui, savait ce qu'il faisait. Ce n'était manifestement pas le cas de Wiener. Lorsque, après sa rupture publique avec l'*establishment* scientifique et militaro-industriel[54] , il se met à déployer une activité d'intellectuel qui va occuper une portion croissante de son temps, c'est un message humaniste qu'il délivre, une philosophie de la technique généreuse et quelque peu naïve, où l'on reconnaît les thèmes de la neutralité de la technique, du mal qu'elle peut faire si on ne la maîtrise pas, de la responsabilité du scientifique et de l'ingénieur, de l'aide décisive que peuvent apporter les machines à l'avènement d'une société rationnelle[55] – bref, un discours de maîtrise, qui fait de la technique un *instrument* que l'on peut et doit dominer, et qui, donc, passe « à côté de son essence ». Tout à fait symboliques à cet égard sont les recherches que Wiener réalise à la même époque, et dont il entretient volontiers ses collègues des conférences Macy, pour mettre au point des prothèses médicales au service des handicapés, en recourant aux techniques de traitement du signal. Nul doute que si beaucoup se méprennent sur la nature philosophique de la cybernétique, c'est parce qu'ils s'arrêtent au discours de Wiener, sans voir qu'il est en contradiction avec le sens de l'œuvre scientifique effectivement accomplie. C'est une raison de plus pour remettre au premier plan la figure de Warren McCulloch[56].

5

Les thèmes cybernétiques : information, totalisation, complexité

Warren McCulloch : « Comme presque tous les Écossais, je tombe amoureux des machines, de certaines machines en particulier, comme les bateaux. »
Margaret Mead : « Mais la machine ne tombe pas amoureuse de vous ! »
McCulloch : « Je n'en suis pas si sûr[1]. »

Ceux qui ont donné de la cybernétique l'image d'une science en rupture avec la physique, qu'ils l'aient fait pour l'en louer ou au contraire pour l'en blâmer, ont généralement présenté les choses ainsi : la physique serait la science de la matière et de l'énergie, la cybernétique celle des *formes.* Cette opposition est irrecevable et elle repose, nous l'avons déjà observé, sur une méconnaissance du rôle essentiel joué par la modélisation mathématique dans la science moderne. La physique est la première à abstraire des formes des phénomènes qu'elle étudie, et à se donner par là même les moyens d'établir des *isomorphismes* entre phénomènes matériels différents, décelant, par exemple, la même structure ondulatoire derrière le mouvement des marées, la vibration d'une corde ou la dynamique d'un électron. Si la cybernétique innove, ce n'est donc pas dans sa démarche modélisante, c'est dans le fait qu'elle étend celle-ci à des domaines nouveaux, que l'on pouvait avant elle juger réfractaires à tout effort de ce type : le système nerveux, les facultés de l'esprit, etc. Cependant, il est probable que la confusion trouve aussi sa source dans l'importance accordée par les cybernéticiens à la question de l'*information.* Or ce serait une erreur d'assimiler forme et information.

Information et physicalisme

Notons d'abord que la notion d'information, dans l'esprit des cybernéticiens, et en tout cas dans celui de Wiener, appartient au domaine de la physique, et plus précisément de la thermodynamique. Dès le symposium de 1946 sur les « Mécanismes téléologiques », Wiener propose de considérer l'information comme étant de l'entropie négative – l'idée avait été avancée par L. Szilard dès 1929, et elle allait être développée de façon rigoureuse par L. Brillouin, en 1956, dans son livre *La Science et la Théorie de l'information* –, et ajoute : « De fait, il n'est guère surprenant que l'entropie et l'information soient de signes opposés : l'information mesure l'ordre et l'entropie mesure

le désordre. Il est en vérité possible de concevoir tout ordre en termes de message[2] ». C'est dire que d'emblée, par le fait même qu'il traite l'information comme une notion physique, Wiener la fait échapper au strict domaine de l'ingénierie des communications, à quoi la limitera Shannon, pour la faire entrer dans celui de l'étude des systèmes organisés, qu'ils soient biologiques, techniques ou sociaux. C'est une *physique de l'information* qu'il prétend établir[3] . C'est la notion d'information qui lui sert à ramener les questions biologiques ou mêmes sociales à des considérations de physique. Ainsi peut-il dire : « La caractéristique la plus importante d'un organisme vivant est son ouverture au monde extérieur. Cela signifie qu'il est doté d'organes de couplage qui lui permettent de recueillir des messages du monde extérieur, lesquels messages décident de sa conduite à venir. Il est instructif de considérer cela à la lumière de la thermodynamique et de la mécanique statistique[4]. »

Le « physicalisme » de Wiener par rapport à la question de l'information apparaît spécialement lors des nombreuses discussions qui portèrent sur les catégories, chères aux cybernéticiens, de « numérique » (*digital*) et d'« analogique ». Les participants aux conférences Macy semblaient avoir beaucoup de mal à percevoir la nature de la distinction, tout en la jugeant fondamentale, puisque beaucoup pensaient que les dispositifs « numériques », de par leur caractère discontinu, étaient davantage porteurs d'information que les dispositifs « analogiques », ou en tout cas qu'ils transmettaient une information plus noble, parce que « codée ». Pitts conjecture que c'est à une étape tardive de l'évolution que les systèmes nerveux sont devenus numériques, « probablement dans le but de traiter de plus grandes quantités d'informations[5] ». Von Neumann distingue entre les messages « codés » parce que discrets (comme l'influx nerveux) et les messages continus, de type hormonal. Il conjecture que les premiers se propagent le long de chemins bien précis et spécialisés, ce. qui n'est pas le cas des seconds[6] . Certains même, tel McCulloch, doutent que l'« analogique » – comprendre le continu – puisse avoir d'autre fonction informationnelle que celle que remplit la différence de potentiel aux bornes d'un poste de radio[7] . Si les discussions sur ce sujet étaient si vives, c'est que l'enjeu en était la sempiternelle question de savoir quelle est l'approche la plus féconde du système nerveux : la « discontinuiste », en termes de réseaux de neurones et d'impulsions nerveuses, ou la « continuiste », en termes de champs électriques et de variables continues, de type chimique et hormonal. Le neurobiologiste Ralph Gerard se faisait régulièrement l'avocat de ces thèses « holistes » et « gestaltistes », allant même jusqu'à conjecturer que le caractère discret de l'influx nerveux n'est qu'une donnée contingente[8].

C'est dans ce contexte que Wiener dut expliquer que l'opposition entre numérique et analogique n'est qu'une distinction de degré et non pas de nature. Il le fit dans des termes très modernes, ceux de la théorie des systèmes dynamiques, dans un langage qui préfigure celui de la théorie des bifurcations, ou des « catastrophes ». Soit un système physique dont les états d'équilibre sont peu nombreux, et les bassins d'attraction[9] – c'est-à-dire les régions de conditions initiales qui conduisent à un même équilibre – bien séparés. La plupart du temps, le système se trouvera en l'un ou l'autre de ses états d'équilibre. La dynamique sous-jacente est continue, mais la phénoménologie est discrète. En général, une petite variation des conditions initiales n'a pas d'effet sur l'équilibre atteint, mais si l'on se trouve en un point de bifurcation, le système saute brusquement, « catastrophiquement », d'un état d'équilibre à un autre, jouant alors le rôle

d'amplificateur. Or, continue Wiener, plus les équilibres sont attractifs, plus le temps que le système passe hors de ses états d'équilibre est faible et plus il se comporte comme un dispositif « numérique ». Un système est donc *plus ou moins* numérique, plus ou moins analogique. Si son identité, et donc le « paysage de ses attracteurs », varie régulièrement en fonction de certains paramètres, on peut même lui faire parcourir toutes les étapes intermédiaires entre le « pur » numérique et le « pur » analogique. Wiener concluait son explication en affirmant la nécessité d'élaborer une « *physique* des dipositifs numériques[10] ».

De façon consistante avec ces développements remarquables, Wiener jugeait vain de prendre parti dans la querelle qui opposait Ralph Gerard à Warren McCulloch, les renvoyant, de fait, dos à dos. Contre le premier, il affirmait que la canalisation des messages le long de sentiers bien précis est une caractéristique essentielle du système nerveux ; contre le second, il conjecturait que les variables continues, d'ordre chimique et hormonal, sont bien porteuses d'information. Pour concilier ces deux positions, il suffit de poser que ces variables « analogiques » modifient continûment les seuils d'excitation des neurones, donc de concevoir un système numérique commandé analogiquement. Il faut cette complexité, pensait Wiener, pour que le système soit capable d'apprendre[11] . L'évolution ultérieure des connaissances lui a donné raison, si l'on en croit cette citation de Jean-Pierre Changeux dans *L'Homme neuronal* : « L'hypothèse d'un *codage chimique* se vérifie, mais celui-ci n'élimine en rien les deux modes, déjà mentionnés, de codages fondés l'un sur la géométrie des connexions, l'autre sur la succession dans le temps des impulsions nerveuses. Il les complète. Il permet d'abord un type de signalisation additionnel qui ne fait pas intervenir la circulation d'impulsions le long des câbles, mais la diffusion « à distance » de signaux chimiques par l'intermédiaire, par exemple, du sang. Ensuite et surtout, il crée une diversité au sein de connexions qui pourraient présenter une géométrie semblable. [...] L'étiquetage chimique diversifie [...]. [Cette diversité] introduit une « combinatoire de signaux », des possibilités de calcul qu'un neurone strictement « électrique » effectuerait différemment ou n'effectuerait même pas du tout[12]. »

Wiener supposait aussi que les messages humoraux, qu'il nommait plaisamment « messages à qui de droit » (*to whom it may concern*), sont le substrat physiologique de l'émotion et de l'affectivité, et croyait possible d'édifier sur cette hypothèse une théorie de l'inconscient qui ne soit pas étrangère à une théorie de la communication[13] . Sur cet exemple de choix, nous voyons une fois encore s'opposer les personnalités de McCulloch et de Wiener. Le premier est l'auteur du modèle de réseau neuronal, mais son idée fixe, pour ne pas dire son fanatisme, lui rend difficile de lui apporter les amendements souhaitables ; ce que nous avons appelé l'éclectisme du second lui permet de se mouler dans la pensée d'un autre, tout en gardant par rapport à elle une distance, et donc une flexibilité salutaires. Dans le groupe des cybernéticiens et de leurs amis, Wiener était sans doute une exception[14].

Dans l'usage qui en est fait tout au long des conférences Macy, la « théorie de l'information » apparaît beaucoup moins comme la clef d'une nouvelle vision du monde que comme un outil familier, un simple *moyen* que l'on utilise à des fins très diverses. Les cybernéticiens apprennent aux autres à s'en servir. Dès la première conférence, le psychanalyste Kubie est sermonné pour parler comme Freud d'« énergie » psychique : il saura désormais distinguer l'information de l'énergie[15] . Plus tard, on lui enseignera à

parler des névroses en termes de réduction de la quantité d'information que le système nerveux peut traiter, pour cause d'indisponibilité de certains de ses circuits[16] . Puisqu'il n'arrive pas à justifier sa thèse selon laquelle la communication symbolique commence avec l'espèce humaine et son langage, on lui apprendra à définir les symboles comme les produits de transformations qui conservent la quantité d'information tout en changeant la nature des objets sur lesquels elle portent[17].

La théorie de l'information sert aussi à apprécier la capacité du système nerveux à reconnaître et à comprendre un même *pattern* abstrait derrière des stimuli différents, obtenus les uns à partir des autres par distorsions réglées. Le psycho-acousticien Joseph Licklider, dont les compétences marient l'ingénierie électrique et la psychologie, rapporte les expériences passionnantes qu'il mène sur « le type et l'importance des distorsions que l'on peut apporter à la parole sans que celle-ci cesse d'être intelligible[18] ». Il s'agit en fait de recherches militaires qui ont commencé pendant la guerre à l'université Harvard, le problème posé étant celui de la bonne compréhension par ses subordonnés des ordres d'un supérieur, donnés et reçus en environnement troublé. L'analyse mathématique que Licklider a conduite pour mesurer l'« intelligibilité » d'un message l'amène à proposer une formule voisine de celle de Shannon. Il s'en autorise pour conjecturer que « l'auditeur reçoit bien de l'information lorsqu'il comprend ce qu'on lui dit[19] ».

La théorie de l'information sert encore à évaluer la redondance d'une langue comme l'anglais et à s'interroger sur son éventuel caractère fonctionnel[20] . Plus prosaïquement, on calcule des quantités d'information à tout bout de champ, fréquemment pour contester ou infirmer une hypothèse. Von Neumann doute ainsi, comme on l'a déjà signalé, que la quantité d'information au niveau neuronal soit suffisante pour rendre compte des capacités de la mémoire humaine, et Pitts lui répond sur le même ton, par un calcul plus complet[21] . Lors de sa conférence au symposium Hixon, McCulloch estime, en citant Wiener, que l'information contenue dans nos chromosomes peut au plus spécifier les connexions entre 10 000 neurones, alors que nous en possédons 1010. Il en déduit que seules les grandes lignes de la structure du cerveau sont génétiquement déterminées, le reste devant être laissé au hasard, c'est-à-dire à l'expérience et à l'apprentissage[22] . Cette conjecture est remarquable, si on la compare aux théories actuelles de l'« enveloppe génétique » et de l'« épigenèse par stabilisation sélective » des neurones et des synapses en développement[23] . McCulloch se réfère à la suggestion de Ramon y Cajal, selon laquelle l'apprentissage correspond à la mise en place de nouvelles connexions. A la neuvième conférence Macy, Henry Quastler présente ses estimations de la « complexité » des organismes, au sens de la quantité d'information qu'ils contiennent. Il conjecture que l'adaptation ne peut être le fin mot de l'évolution, parce qu'elle est incapable d'expliquer la *complexification* des êtres vivants[24].

Est-ce à dire que Raymond Ruyer est fondé à écrire, dans son livre de 1954, *La Cybernétique et l'Origine de l'information*, que « les théoriciens aussi bien que les techniciens de la cybernétique sont au fond peu intéressés par tout ce qui n'est pas problème technique immédiat » et qu'ils « estiment qu'ils n'ont pas à être plus philosophes que les psychologues behaviouristes[25] » ? Cette appréciation me paraît tout à fait injuste et malvenue. Ruyer reproche aux cybernéticiens de ne jamais poser le problème de l'*origine* de l'information, et cela parce qu'ils seraient bien incapables d'y faire face, prisonniers qu'ils sont de leurs postulats mécanistes[26] . Il est bien vrai que cette question ne préoccupe guère les cybernéticiens, mais c'est précisément parce que, dès le

départ, pour ce qui est de Wiener et de ceux qui le suivent en tout cas, ils ont fait de l'information une grandeur *physique*, l'arrachant au domaine des transmissions de signaux entre humains. Si tout organisme est environné d'informations, c'est tout simplement qu'il y a partout autour de lui de l'organisation, et que celle-ci, du fait même de sa différenciation, *contient* de l'information. L'information est dans la nature, et son existence est donc indépendante de l'activité de ces donneurs de sens que sont les interprètes humains. C'est sur des conceptions similaires que, dans la philosophie cognitive contemporaine, le philosophe américain Fred Dretske, ingénieur de formation, a proposé une théorie naturaliste et physicaliste de l'« intentionnalité » qui, de toutes les doctrines aujourd'hui en présence, est certainement celle qui pousse le plus loin le projet de Quine d'une « épistémologie naturalisée »[27] – mais qui donc, aussi, mène à son comble la méprise du même au sujet de l'intentionnalité brentanienne.

Le problème qui préoccupe les cybernéticiens serait plutôt l'inverse : c'est l'énorme « gaspillage » d'information dont se paie le passage de la sensation à la perception. McCulloch observe que l'œil ne transmet au cerveau que le centième de l'information qu'il reçoit. Il y a une raison à cela, conjecture-t-il. Il faut cet énorme excès d'information pour nous assurer que tel événement qui affecte tel point de nos organes des sens n'est pas dû à un bruit, mais correspond à un objet extérieur. Nous payons en information la certitude qu'il existe un monde au dehors de nous[28] . A la septième conférence Macy, lors de la discussion qui suit l'exposé de Shannon sur la redondance de la langue anglaise, Pitts et Stroud évoquent l'« immense perte d'information » qui se produit entre nos organes des sens et notre « computer » mental, et conjecturent que cette perte est en fait programmée. Pitts souligne le rôle des « lois de la nature », dont l'existence même traduit la très grande redondance de celle-ci[29]. C'est cette redondance qui fait office de réservoir d'information. Sur ce point aussi, la filiation entre la cybernétique et un Fred Dretske est évidente.

Les cybernéticiens n'étaient ni des techniciens bornés, ni les champions d'une science nouvelle bien décidés à faire table rase de la science « orthodoxe ». Ces deux visions sont aussi fausses l'une que l'autre. C'étaient des scientifiques qui s'exerçaient au maniement d'instruments conceptuels nouveaux, tout en réfléchissant constamment à l'emploi de ces outils. Configuration assez rare dans l'univers scientifique, certes, mais qui n'autorise nullement à parler d'une quelconque « rupture épistémologique ». La façon dont furent discutés les rapports entre information, forme, hasard et sens, illustre bien ce que nous voulons dire.

L'information : entre la forme, le hasard et le sens

Nous sommes à la septième conférence Macy, en 1950. Claude Shannon vient de présenter les résultats d'expériences qu'il a récemment conduites pour évaluer la redondance de l'anglais écrit. Il s'agit de faire deviner à un sujet un texte qu'il ne connaît pas, lettre après lettre. Shannon a pris soin de rappeler d'emblée que sa notion de quantité d'information est définie indépendamment du sens du message. Il est fascinant de voir comment la discusion qui suit son exposé est progressivement amenée à ignorer cette exclusion fondatrice. Le retour du sens est inévitable, il s'accompagne d'une interrogation majeure sur la possibilité de dire sans arbitraire ce qui, dans un message, est signal et ce

qui est bruit[30] . La tension est nette entre ceux qui prennent le point de vue de l'ingénieur, c'est-à-dire de celui qui décide souverainement de ce qui est signal et de ce qui est bruit, et ceux qui prennent le point de vue de l'organisation ou de l'organisme étudiés, et par rapport auxquels une perturbation peut devenir sens, indépendamment de la liberté de l'observateur d'attribuer et d'imposer des significations au système observé. Bateson évoque le cas des fautes de frappe qui produisent des « distorsions signifiantes ». Pitts lui réplique aussitôt : « On peut les corriger extrêmement facilement[31]. »

L'idée se fait jour que hasard et sens sont peut-être les deux faces d'une même médaille, et que l'organisation est un mixte de redondance et de variété, de régularités et d'anomalies. Licklider prend le cas d'une communication entre deux personnes. Marquis vient de lancer une boutade : « La meilleure communication, c'est lorsque vous dites précisément ce à quoi s'attend votre interlocuteur. » Licklider répond que non, et que l'optimum doit se situer entre deux extrêmes. Si la « corrélation » est nulle entre le locuteur et l'auditeur, celui-ci n'a aucune attente et ne comprend rien. Si elle vaut un, il n'a pas besoin d'écouter. Heinz von Foerster enchaîne sur le cas d'une langue. Il doit y avoir un optimum de redondance, suggère-t-il. Plus la redondance est faible, plus la langue peut transmettre d'information ; mais plus elle est forte, mieux elle est structurée. D'autres vont vite oublier cette réflexion, et juger que moins une langue est redondante (par exemple l'anglais !), plus elle est éloignée de l'« état primitif »[32].

Cette idée, selon laquelle il n'y a d'organisation que dans un entre-deux entre l'ordre et le désordre, sera au cœur de la seconde cybernétique, avec le principe d'« ordre par le bruit » (*order from noise*) formulé par Heinz von Foerster en 1960[33] , puis de la théorie de l'auto-organisation développée par Henri Atlan[34] . Mais dès 1951, dans sa « Note by the Editors » publiée dans les Actes de la huitième conférence Macy, von Foerster écrit : « l'information peut être considérée comme de l'ordre arraché au désordre » (*order wrenched from disorder*[35]). Ce thème imprègne toute l'œuvre de Wiener, comme en témoigne la citation suivante, extraite de son autobiographie tardive *I Am a Mathematician* : « Nous nageons à contre-courant d'un immense torrent de désorganisation qui tend à réduire toutes choses à la mort thermique de l'équilibre et de l'indifférenciation décrite par le second principe de la thermodynamique. Ce que Maxwell, Boltzmann et Gibbs entendaient par mort thermique en physique trouve son équivalent dans l'éthique de Kierkegaard, pour qui nous vivons dans un univers moral chaotique. Dans ces circonstances, notre principale obligation est d'établir des enclaves arbitraires d'ordre et de système. Ces enclaves ne persévéreront pas dans leur être indéfiniment une fois que nous les aurons établies. A l'instar de la Reine de cœur, nous ne pouvons rester là où nous sommes que si nous courons aussi vite que nous le pouvons.

« Nous ne nous battons pas en vue d'une victoire définitive pour un avenir indéfini. La plus grande victoire possible, c'est de continuer à être, et d'avoir été. Aucune défaite ne nous retirera le succès d'avoir existé un certain temps dans un univers qui nous semble indifférent.

« Il n'y a ici aucun défaitisme, plutôt un sens du tragique par rapport à un monde dans lequel la nécessité prend la forme de la disparition inévitable de toute différenciation. L'affirmation de notre nature propre et l'effort de construire une enclave d'organisation face à la tendance de la nature à tout submerger dans le désordre envahissant, c'est notre insolence devant les dieux, et devant la nécessité de fer qu'ils nous imposent. La tragédie est ici, mais aussi la gloire[36]. »

Les représentants du mouvement « personnalité et culture » buvaient du petit lait en entendant les propos lyriques de Wiener. A la septième conférence Macy, Lawrence Frank s'emploie sentencieusement à lui emboîter le pas, expliquant que toute culture crée un monde à partir du seul bruit de fond des événements, sélectionnant dans ce chaos certains signaux qu'elle traite comme des messages, en choisissant de leur attribuer du sens[37].

Dans son très bel ouvrage, *Du mode d'existence des objets techniques*, Gilbert Simondon appuie sa réflexion sur une abondante littérature cybernétique, dont – la chose est trop rare pour ne pas être notée – les Actes des conférences Macy. Prenant ses distances, il écrit : « Ce qui risque de rendre le travail de la cybernétique partiellement inefficace comme étude interscientifique (telle est pourtant la fin que Norbert Wiener assigne à sa recherche), c'est le postulat initial de l'identité des êtres vivants et des objets techniques autorégulés[38] . » Il est d'autant plus remarquable qu'il crédite la cybernétique de l'idée suivante : « L'information est [...] à mi-chemin entre le hasard pur et la régularité absolue. On peut dire que la forme, conçue comme régularité absolue, tant spatiale que temporelle, n'est pas une information mais une condition d'information ; elle est ce qui accueille l'information, l'*a priori* qui reçoit l'information [...]. Mais l'information n'est pas de la forme, ni un ensemble de formes, elle est la variabilité des formes, l'apport d'une variation par rapport à une forme. Elle est l'imprévisibilité d'une variation de forme, non la pure imprévisibilité de toute variation[39]. »

On doit cependant à la vérité de reconnaître que, en dehors de Wiener, les cybernéticiens ne se montraient pas très réceptifs à l'idée que le hasard peut, en certaines circonstances, produire du sens. Le Britannique Donald MacKay, qui avait participé, avec Grey Walter et Ross Ashby, au « Ratio Club » qu'animait Alan Turing, en sut quelque chose lorsque McCulloch l'invita à participer à la huitième conférence, en 1951. Ayant exposé sa conception d'un automate qui serait doué de la capacité de faire des inférences inductives en ayant recours à des stratégies aléatoires, il se fit sévèrement critiquer par Leonard Savage, pourtant statisticien de son état. Celui-ci développa longuement l'idée que l'incorporation de hasard dans un mécanisme ne peut en rien l'aider à mimer un comportement humain, et en tout cas pas à accroître son efficacité dans la résolution de problèmes[40] . Mais c'est en réaction au choc que représenta la participation de Ross Ashby à la neuvième conférence que les préventions des cybernéticiens à l'égard du hasard se révélèrent le mieux[41]. Nous y reviendrons.

Claude Shannon avait bâti une théorie de l'information qui excluait toute référence à la signification. Il était inévitable que d'autres tentent de construire une théorie complémentaire dont l'objet serait l'information sémantique. Deux de ces tentatives furent présentées et discutées aux conférences Macy. Celle de Donald MacKay à la huitième, et celle de Rudolf Carnap et de son disciple Yehoshua Bar-Hillel à la dixième[42] . Mais c'est la théorie de Shannon que les psychologues expérimentaux proches des cybernéticiens préféraient néanmoins prendre comme cadre conceptuel. Ils tombaient inévitablement alors sous le feu des critiques de leurs collègues gestaltistes ou « holistes », qui ne manquaient pas de leur prouver que les significations qu'ils pensaient avoir neutralisées revenaient interférer, quoi qu'ils fissent, avec leurs plans d'expériences. A la sixième conférence, John Stroud fit le compte rendu des expériences qu'il menait pour la Marine et dont l'objectif était d'évaluer la quantité maximale d'information (au sens de Shannon) qu'un homme est capable d'absorber par unité de temps. Il fut facile à

Klüver de montrer que ces résultats n'avaient ni grand sens ni grand intérêt, car cette capacité varie énormément suivant que l'on présente au sujet une succession de symboles sans signification ou des séquences faisant globalement sens. Après que von Foerster eut présenté au groupe la théorie quantique de la mémoire dont il était l'auteur et qui avait retenu l'attention de McCulloch, Klüver revint à la charge, disant : « Le non-sens absolu, c'est-à-dire un matériau totalement dépourvu de signification et ne donnant prise à aucune association, semble être demeuré un idéal qui n'a jamais été réalisé[43]. » Le point de vue « holiste », dans ces discussions, était toujours dirigé *contre* l'esprit cybernétique.

Entre ceux qui, comme Shannon, Stroud, Licklider et d'autres, croyaient *pratiquement* possible de faire abstraction du sens, et ceux qui tenaient cela pour impossible, le conflit ne portait pas seulement sur la nature de l'information. C'étaient deux conceptions de la rationalité qui s'opposaient. La ligne de clivage, ici, ne recoupe pas celles que nous avons déjà repérées : ce ne sont pas les cybernéticiens contre les « holistes » et autres gestaltistes, ni les « aprioristes » contre les psychologisants, ni le style de McCulloch contre celui de Wiener. Il s'agit de deux attitudes opposées par rapport à la rationalité économique qui, à l'époque, fait deux avancées décisives dans la voie de l'axiomatisation et de la mécanisation, avec la publication de l'ouvrage de John von Neuman et Oskar Morgenstern, *Theory of Games and Economie Behavior*, en 1944, et la mise au point de celui de Leonard Savage, *Foundations of Statistics*, qui allait paraître en 1954. La théorie des jeux et la théorie de la décision, dès leur naissance en tant que théories formalisées, auront donc participé à l'éclosion de la cybernétique et, au-delà, des sciences cognitives. L'effort de synthèse le plus ambitieux reste à ce jour celui de Herbert Simon, pionnier tout à la fois de la recherche en gestion (*administrative behavior*) et de l'intelligence artificielle. Mais à l'époque de la cybernétique, c'est sur le mode conflictuel que les échanges ont lieu. Opposé à cette forme d'esprit si particulière de l'économiste, qui mêle souci obsessionnel pour la rigueur logique et représentation délibérément appauvrie des relations humaines – forme d'esprit qui, aux conférences Macy, est représentée par von Neumann, Savage, Lazarsfeld, dans une une certaine mesure Pitts, et les ingénieurs de la communication comme Shannon, Stroud et Licklider –, on trouve un groupe qui réunit, une fois n'est pas coutume, Wiener, McCulloch et Bateson. McCulloch, en particulier, ne perdait jamais une occasion de dire tout le mal qu'il pensait de la théorie de l'utilité, qui aplatit toutes les composantes des désirs et des choix le long d'une dimension unique. Son sang ne fait qu'un tour lorsque, à la huitième conférence, Savage risque ce trait d'humour propre aux économistes, qui n'hésitent pas à dévoiler l'obscène : « La valeur de l'information ? – mais ce n'est rien d'autre que sa valeur *cash*[44] ! » McCulloch était très attaché à la théorie de l'« anomalie axiologique » qu'il avait mise au point en 1945. Il avait réussi à construire un réseau de six neurones qui incarnait de façon cohérente ce qui, pour la théorie du choix rationnel, est incohérence : la circularité des préférences – un agent préférant une option A à une option B, B à C, et C à A. Il n'y a pas d'échelle commune aux valeurs, le bien n'est pas réductible à une commune mesure[45] . Bateson et Wiener, pour leur part, s'employaient à critiquer la théorie des jeux au nom d'une version approximative de la théorie des types logiques de Bertrand Russell : on peut gagner une bataille et perdre la guerre ; on peut gagner la guerre et perdre son âme[46].

Coopération et cognition

La rencontre des thèmes de l'information et de la rationalité se fit à la huitième conférence, à l'occasion de l'étonnante discussion qui suivit la communication du psychosociologue Alex Bavelas. Bavelas avait été l'élève du gestaltiste Kurt Lewin, l'un des membres réguliers du groupe cybernétique depuis la première conférence, mais qui devait mourir peu de temps avant la troisième. Dans son livre *The Cybernetics Group*, Steve Heims décrit plaisamment l'itinéraire qui mène de Kurt Lewin, élève de Husserl et de Stumpf, membre de l'Institut psychologique de Berlin, à son élève américain Alex Bavelas, par l'expression à double sens « *Gestalten* Go to Bits » (« Quand les *Gestalten* partent en morceaux et se fragmentent en *bits* d'information »)[47]. Arrivé en 1933 aux États-Unis comme réfugié, Lewin eut une influence considérable sur la psychologie américaine des années trente, quarante et cinquante. Comme Alfréd Schutz, il tenta d'abord de bâtir une science sociale, plus précisément, une psychologie sociale, sur les bases de la phénoménologie husserlienne. Au concept de *Lebenswelt* chez Schutz correspond celui de *life space* chez Lewin. Mais ce dernier avait l'ambition de faire de la psychologie une science mathématique et expérimentale à l'égal de la physique, et c'est, comme son mentor Wolfgang Köhler, dans le concept physique de champ, ainsi que dans la topologie mathématique qui était encore dans l'enfance, qu'il plaçait les plus grands espoirs. Les concepts de « dynamique de groupe » et de « recherche-action », qui étaient, l'un et l'autre, promis à un brillant avenir, sont l'œuvre du génie de Lewin. Dès 1946, il découvrit les concepts de base de la cybernétique et de la théorie des jeux grâce aux conférences Macy, et chercha aussitôt à les appliquer à ses recherches sur la psychosociologie des groupes, par exemple les mécanismes d'interaction entre un chef et une foule. Sa mort prématurée ne permet pas de dire ce qu'eût été, comme fruit de ses efforts, une science sociale mariant le souci, hérité de la phénoménologie, des totalités et les outils cybernétiques.

Alex Bavelas n'avait pas la stature de son maître, mais il poursuivait la tradition inaugurée par celui-ci en menant des recherches-actions en laboratoire et en milieu industriel. Il s'agissait d'analyser des situations de coopération, de comprendre les obstacles qui s'opposaient à une meilleure productivité et de concevoir les moyens de les lever en repensant les systèmes de communication. Un type de recherche, on le voit, lui aussi destiné à faire école. A la huitième conférence, donc, Bavelas présenta au groupe certains de ses travaux en usant d'un appareil conceptuel qui ne devait presque plus rien aux « espaces de vie » de son maître Lewin et presque tout à la théorie de l'information et aux concepts cybernétiques. Le problème de base est de faire réaliser à un nombre réduit de personnes une tâche collective qui exige la coopération de tous. Des informations doivent être échangées entre les partenaires, mais le dispositif physique de l'expérience en laboratoire, ou de la situation effective en milieu industriel, contraint les canaux de communication disponibles, isolant plus ou moins chacun et, en tout cas, limitant le nombre de ses interlocuteurs directs.

Bavelas décrivit plusieurs de ces expériences et il en vint à la suivante, une des plus simples. Cinq personnes, totalement isolées les unes des autres, ont chacune à inscrire sur un morceau de papier un nombre entier compris entre 0 et 5, le total devant être égal à 17. Première version : après chaque coup, le meneur de jeu annonce à tous le total réalisé. On recommence jusqu'à l'obtention de la bonne valeur. Seconde version : le meneur de jeu dit simplement : « Vous n'y êtes pas, recommencez », jusqu'à ce que le bon total soit atteint. Or les expériences de Bavelas montraient sans ambiguïté que tous les groupes

étudiés convergeaient plus vite dans le second cas que dans le premier.

De tels apparents paradoxes ont le mérite de susciter des réactions qui révèlent des attitudes profondes et des visions du monde contrastées. La discussion qui suit l'exposé de Bavelas partage en gros les participants en deux camps (malheureusement, Wiener, von Neumann et Bateson sont absents). D'un côté, Shannon, Savage et Bigelow parlent, tout simplement, d'*irrationalité.* La seule rationalité concevable est celle de l'algorithme. Mettez une machine à résoudre le problème, il est inconcevable qu'elle fasse moins bien si on lui donne une information supplémentaire. La valeur d'une information ne peut jamais être négative. Oui, mais *cinq* machines ? demande toutefois Savage. En face, Bavelas, MacKay, Kubie, Klüver et quelques autres cherchent à comprendre. Kubie pense aux déterminations inconscientes, au « sens de l'insensé », et n'écarte pas la parapsychologie. Bavelas suggère une logique de la spécularité. Lorsque le meneur de jeu donne le total obtenu, chacun raisonne en se mettant à la place des autres, et perçoit que les autres font de même à son égard. D'où une indétermination accrue, que les groupes étudiés cherchent d'ailleurs à réduire en élaborant des théories sur la répartition des rôles dans le rééquilibrage de l'opération. Un ensemble humain ne reste jamais longtemps sans structuration sociale. (Ces expériences ont depuis été répétées de nombreuses fois, et il semble que les membres de sociétés hiérarchiques, comme la société japonaise, y réussissent mieux que d'autres. Dans le groupe d'individus sous investigation, chaque membre est capable de repérer celui d'entre eux qui, dans la société globale, occupe la position hiérarchique la plus élevée. Sa position éminente est connue de tous, et chacun sait que les autres la connaissent. Le groupe converge alors très vite vers la stratégie suivante : les quatre membres non distingués choisissent 3, le membre distingué choisissant 5.)

MacKay se veut œcuménique et tente de raccorder l'interprétation de Bavelas à la théorie de l'information de Shannon. Mais les cybernéticiens purs et durs restent sourds à ces considérations. On n'a jamais vu une machine s'identifier à une autre, ni des inégalités se mettre en place spontanément au sein d'une société d'automates. D'une discussion confuse sur la théorie des jeux de von Neumann il ressort qu'il n'y a rien à attendre d'elle pour éclairer des situations de ce genre, qui sont des situations de coopération et non de conflit.

On voit ici à l'œuvre, sur cette question de la rationalité, ce qu'a été une certaine fermeture de l'esprit cybernétique, et combien le progrès de la réflexion dans les domaines mêmes qui étaient les siens en a été retardé. Il fallut en effet attendre les années soixante, avec la publication de l'ouvrage du théoricien des jeux Thomas Schelling, *The Strategy of Conflict*, en 1960, et de celui du philosophe analytique David K. Lewis, *Convention*, en 1969[48] , pour que l'on comprenne que les « jeux de coordination », s'ils ne présentent certes qu'un intérêt mathématique médiocre, posent des problèmes passionnants dans l'ordre cognitif. Ils mettent en jeu la capacité des agents à se mettre à la place des autres et à voir le monde, y compris eux-mêmes, de leur point de vue. Cette capacité pouvant se refléter ainsi dans l'esprit des autres jusqu'à l'infini, on débouche sur un concept, le « savoir public » (*common knowledge*), caractérisé comme spécularité infinie. Ce concept joue aujourd'hui un rôle clé dans nombre de disciplines, de l'économie à l'intelligence artificielle, de la réflexion sur les fondements de la théorie des jeux et du choix rationnel à la philosophie politique[49]. Il n'est pas exagéré de dire que si la théorie des jeux a connu à partir des années soixante-dix une seconde jeunesse, et si elle a

pu devenir, plus qu'une théorie au domaine limité, une des principales méthodes de pensée dans la philosophie analytique et dans les sciences sociales, c'est à la réflexion sur les jeux de coordination qu'elle le doit. Ce fut donc un rendez-vous manqué de plus pour la cybernétique.

Je conjecture que si, de tous les cybernéticiens au sens large, ce fut Donald MacKay qui se montra le plus réceptif à ce que tentait d'exprimer Alex Bavelas, la raison en revient à son appartenance au cercle des proches de Turing. Nous sommes en 1951, et celui-ci vient de publier l'article où il présente le « jeu de l'imitation », dont nous avons parlé au chapitre premier. Si les cybernéticiens s'étaient pénétrés du message de cet article, à savoir que penser c'est simuler, eux qui, par ailleurs, n'écartaient pas la possibilité qu'une machine pût penser – non pas qu'ils la dotassent de facultés humaines, mais parce qu'ils croyaient que la pensée est un mécanisme – n'auraient pas pris la même attitude vis-à-vis des travaux de Bavelas.

Savage, cependant, sauva l'honneur en concluant la discussion par un singulier aveu. Depuis que nous nous réunissons, dit-il en substance, nous avons été déchirés entre deux espoirs contradictoires. Le premier, nourri par le « théorème » de Turing (*sic* !), c'est que nous trouverons toujours une machine capable de faire ce que fait l'homme, quel que soit le comportement spécifié ; le second, c'est que nous découvririons enfin quelque chose dans l'homme que la machine ne saurait reproduire[50]

Les totalités cybernétiques

Parce qu'elle s'est instituée comme la science des « mécanismes téléologiques », et qu'elle a pris les êtres finalisés pour objets d'étude, on considère souvent la cybernétique comme le fruit de l'union d'une philosophie « holiste » et de la science moderne. Cette image est aussi fausse que les autres, et rien ne vaut la lecture des Actes des conférences Macy pour la démystifier.

Les critiques les plus sévères de la cybernétique ne cachent pas leur fascination pour les automates qu'elle a conçus. Ainsi, Raymond Ruyer, dans son livre *La Cybernétique et l'Origine de l'information*, peut-il écrire, à propos des travaux de Ross Ashby, dont nous parlerons à la fin de ce livre parce qu'ils ouvrent le passage vers la seconde cybernétique : « Il faut reconnaître que l'homéostat d'Ashby apporte un élément vraiment sensationnel dans le débat. […] [Il] n'est évidemment qu'une machine, et pourtant, il agit comme un tout, qui serait indépendant de ses parties, de leurs lésions ou servitudes accidentelles. On imagine aisément les développements philosophiques d'un organiciste, disciple de Goldstein, sur les performances de la machine d'Ashby, si on lui jouait le mauvais tour de les lui donner à interpréter en lui cachant qu'il s'agit d'une machine. […] Nous ne prétendons certes pas que la théorie mécaniste d'Ashby soit vraie au fond, mais nous reconnaissons que la cybernétique marque là un point contre les théories antimécanistes[51] . » Dans *L'Automate et ses mobiles*, Jean-Claude Beaune, pour sa part, dit des « petites monstres cybernétiques » qu'ils représentent « un acte philosophique de portée considérable ». « Leur qualité cybernétique transforme leur facticité spectaculaire en logique de l'apparence. Ce sont des métaphores, des paradigmes philosophico-techniques[52]. »

Dans un deuxième temps, les critiques se ressaisissent – « nous avons failli nous y faire

prendre » – et identifient le véritable projet cybernétique : produire, comme l'art, des simulacres, c'est-à-dire des imitations d'imitation, des modèles de modèle. La totalisation cybernétique n'est pas holiste, c'est un holisme artificiel, une simulation de holisme qui agit comme démystification. On ne saurait donc reprocher aux cybernéticiens de tromper leur monde puisqu'ils sont les premiers, après coup, à reconnaître le tour qu'ils nous ont joué. Gilbert Simondon a bien compris la nature émancipatrice de cet exercice. « La cybernétique […] libère l'homme du prestige inconditionnel de l'idée de finalité », écrit-il dans *Du mode d'existence des objets techniques*. « Alors que l'évocation d'une fin supérieure, et de l'ordre qui réalise cette fin, est considérée comme terme dernier d'une requête de justification, parce que la vie est confondue avec la finalité, à une époque où les schèmes techniques ne sont que des schèmes de causalité, l'introduction dans la pensée des schèmes technologiques de finalité joue un rôle cathartique. Ce dont il y a technique ne peut être une justification dernière[53] . » Appliqué au social, le projet cybernétique prolonge l'œuvre de Rousseau qui n'est libératrice que dans la mesure précisément où son holisme est une contrefaçon, un artefact, puisque le tout social y est reconstitué sur la base de la conscience et de la volonté des sociétaires individuels. Simondon écrit : « … L'homme se libère de sa situation d'être asservi par la finalité du tout en apprenant à faire de la finalité, à organiser un tout finalisé qu'il juge et apprécie, pour n'avoir pas à subir passivement une intégration de fait. La cybernétique […] libère l'homme de la fermeture contraignante de l'organisation en le rendant capable de juger cette organisation, au lieu de la subir en la vénérant et en la respectant parce qu'il n'est pas capable de la penser ou de la constituer. L'homme dépasse l'asservissement en organisant consciemment la finalité[54]… »

Les totalités cybernétiques sont toujours des totalités artificielles, dans lesquelles les parties sont antérieures au tout ; des totalités nominales, que seule la conscience organisatrice d'un tiers, en l'occurrence le cybernéticien, vient achever en les percevant et les concevant[55] ; des agrégats dont l'unité est accidentelle, et jamais des monades dont l'unité soit substantielle, pour reprendre la célèbre distinction de Leibniz. Le problème de la cybernétique, si l'on peut dire, c'est qu'il n'y a pas pour elle d'autre forme de totalité concevable. Lorsque les cybernéticiens disent, et ils ne s'en privent pas, à propos des écosystèmes, de l'homéostat d'Ashby, des sociétés humaines, des réseaux d'automates, du rat de Shannon, des colonies de fourmis ou des organismes vivants, qu'ils apprennent, qu'ils choisissent, qu'ils résolvent, qu'ils se souviennent, qu'ils essaient ou qu'ils se trompent, cela n'a pour eux ni plus ni moins de sens que lorsqu'ils disent la même chose d'un cerveau ou d'un esprit humain[56]. Cela n'a pas moins de sens, car l'esprit n'est que l'ensemble des propriétés d'un réseau de neurones interconnectés, et il n'y a pas de différence d'essence entre ces propriétés et celles de n'importe quel système de n'importe quels éléments interconnectés d'une façon similaire ; cela n'a pas plus de sens, car la colonie de fourmis n'est pas plus un superorganisme en surplomb par rapport aux fourmis individuelles que l'esprit n'est une entité transcendante par rapport au réseau de neurones. Que Bateson choisisse de considérer toutes ces totalités comme relevant de l'esprit (*mind*) ou que McCulloch y voie le fondement d'un matérialisme renouvelé est ici tout à fait secondaire. Qu'il se dise spiritualiste ou matérialiste, ce monisme intransigeant se donne pour universel et ne connaît nulle part de coupure ni de différence.

Nous avons vu, au chapitre précédent, comment la philosophie de l'esprit, contrainte par sa conception psycho-linguistique de la représentation, en avait été conduite à se

constituer en un monisme matérialiste non réductionniste (les deux variantes principales étant le fonctionnalisme et le monisme anomal de Davidson). Rares sont donc aujourd'hui les philosophes cognitifs qui défendent, comme les premiers cybernéticiens, un monisme réductionniste, c'est-à-dire la thèse de l'identité entre le cerveau et l'esprit. Le cerveau est identique à l'esprit, parce que l'un et l'autre ne sont que de l'organisation. Il en existe, cependant, mais la pente est glissante qui fait passer de la thèse de l'identité à la thèse « éliminationniste », selon laquelle les catégories de l'esprit ne désignent tout simplement rien, ainsi que l'affirment, aux États-Unis, les « neurophilosophes » Churchland, mari et femme[57] . La glissade dont je parle est allégrement faite par Jean-Pierre Changeux dans *L'Homme neuronal.* Il écrit d'abord, parlant du réseau de neurones : « Ces enchaînements et emboîtements, ces "toiles d'araignée", ce système de régulations fonctionneront *comme un tout.* Doit-on dire que la conscience "émerge" de tout cela ? Oui, si l'on prend le mot "émerger" au pied de la lettre, comme lorsqu'on dit que l'iceberg émerge de l'eau. Mais il nous suffit de dire que la conscience *est* ce système de régulations en fonctionnement. » Mais c'est pour ajouter aussitôt : « L'homme n'a dès lors plus rien à faire de l'"Esprit", il lui suffit d'être un Homme Neuronal[58]. »

Or, très tôt dans sa carrière, la cybernétique devait se trouver confrontée à une autre conception des totalités organisées, plus riche et en un sens dirigée contre elle. Cette confrontation eut lieu au symposium Hixon, en 1948, et elle mit aux prises, d'un côté, Warren McCulloch, de l'autre un ensemble de chercheurs composé des neurophysiologistes Karl Lashley et Ralph Gerard, des psychologues Wolfgang Köhler et Heinrich Klüver, et surtout de l'embryologiste Paul Weiss. John von Neumann était également présent, il dut lui aussi subir les assauts de Weiss, mais il faut surtout retenir de ses interventions la critique que lui-même adressa à McCulloch, dont nous avons déjà parlé et sur laquelle nous reviendrons. La conception présentée par Weiss nous intéresse au plus haut point parce qu'elle annonce dans une certaine mesure les théories de l'autopoiése et de l'autonomie qui seront développées par l'école chilienne de l'auto-organisation (Maturana et Varela)[59]. Du point de vue de l'histoire des idées qui est ici le nôtre, il est particulièrement piquant de voir la cybernétique mise en accusation par un mouvement de pensée que sa descendance contribuera à alimenter.

Quelques mots sur Paul Weiss et l'embryologie sont ici appropriés[60]. J'ai raconté ci-dessus la rencontre ratée entre la cybernétique et ce qui allait devenir la biologie moléculaire, ou comment la physique qui, sous l'impulsion du *What is Life* ? de Schrödinger (1945), s'apprêtait à conquérir la biologie en donnant de l'ordre du vivant une interprétation mécanique, avait snobé la cybernétique. J'ai insisté sur le fait qu'on ne saurait inférer de ce rendez-vous manqué qu'il n'y avait pas une profonde affinité entre les deux démarches. La suite de l'histoire l'a prouvé amplement, puisque c'est grâce à une métaphore cybernétique et informatique, celle du « programme génétique », que la biologie moléculaire a pu se constituer. Or, dans cet immédiat après-guerre, il existe une science du vivant qui, après avoir connu des moments glorieux avant guerre (Hans Spemann obtient le prix Nobel pour ses travaux du début des années vingt sur le triton), va engager, mais vainement, la résistance contre la mécanisation de la biologie qui se met en place : l'embryologie.

Nouveau paradoxe : alors que les physiciens devenus biologistes pensent l'ordre du vivant en termes de rupture, ou d'arrachement au désordre moléculaire qui sert de toile de fond aux lois physiques, les embryologistes tentent de concevoir l'ordre biologique

comme émergeant de ce même désordre, sans aucune solution de continuité. Les premiers, à la suite de Delbrück et de Schrödinger, croient que le vivant est soumis à des lois propres, radicalement différentes des lois physiques. (En 1932, Niels Bohr considérait déjà que le vivant est *irréductible*, à la manière du quantum d'action en mécanique quantique[61].) L'ordre du vivant échappe selon eux à la juridiction de la thermodynamique, puisque c'est un ordre mécanique, tout entier codé dans le matériel héréditaire. (Paradoxe supplémentaire, la biologie moléculaire naissante est bien plus portée que la cybernétique, on l'a vu, à faire de l'information une entité qui échappe à l'ordre des lois physiques.)

A la question : « La science peut-elle rendre compte de l'énigme de la vie, cette anomalie par rapport au mouvement général vers le désordre qui caractérise les phénomènes physiques ? », les embryologistes apportent en revanche une réponse qui est beaucoup plus conforme à la démarche scientifique traditionnelle. L'embryologie analyse expérimentalement les processus physicochimiques responsables du développement et de la différenciation de l'embryon. Elle ne découvre dans ces processus aucune singularité ou anomalie par rapport à ceux qui constituent les objets physiques ordinaires, inorganiques. Elle en infère que la singularité de son objet, cette capacité étonnante de complexification, doit être imputée à une organisation particulière – et plus précisément, pour user d'un terme qui, dès les années cinquante, a acquis un statut technique dans la communauté des embryologistes, une « auto-organisation »[62] . Isabelle Stengers avance à ce propos une conjecture intéressante, que n'infirme pas Ilya Prigogine. Si ce dernier, en posant les bases d'une thermodynamique des processus irréversibles loin de l'équilibre, plus connue sous le nom de théorie des « structures dissipatives », a pu parler de l'« auto-organisation » de la matière dès les années soixante, c'est à sa fréquentation des embryologistes de l'université de Bruxelles qu'il l'a dû. Du côté de l'embryologie scientifique comme de celui de la thermodynamique, la conviction était forte que, entre l'entropie des processus chimiques et la complexité du vivant, le passage existait, et que sa découverte ne requérait aucunement la mobilisation de catégories empruntées à la régulation machinique ou à la théorie de l'information[63].

Né en 1898, Paul Weiss est, après la guerre, l'un des représentants les plus en vue de cette embryologie scientifique dont les années sont désormais comptées. Théoricien de la biologie, il va, face à la biologie moléculaire montante, s'opposer au réductionnisme et au déterminisme de l'information génétique, ainsi qu'à la conquête de la biologie par la cybernétique. Cette citation tardive, puisqu'elle date de 1971, exprime bien ce qu'aura été le sens de sa critique tout au long de sa carrière : quand on a décrit l'ordre biologique en termes de mécanismes et de tranferts d'information, on n'a rien expliqué du tout, il reste tout à faire : « Ainsi susbsiste la question de savoir comment un fouillis d'activités moléculaires peut conduire à un ordre global et intégré ; et comment le comportement imprécis et variable des cellules individuelles conduit à des organes qui sont infiniment plus semblables parmi les membres d'une même espèce que ne le sont les processus détaillés qui leur donnent naissance à travers la morphogenèse [...]. C'est quand on aborde ces problèmes que s'écroule le concept de « transfert d'information », exactement comme le train qui, à cause d'une voie ferrée interrompue, s'enlise dans un désert sableux avant d'atteindre sa destination ; il ne pourrait parvenir au but que si, au-delà du guidage préréglé que constituent les rails, un système de libre navigation prenait la relève. Autrement dit, ce qui nous apparaît parfois comme un espace non structuré n'est pas forcément un désert, mais peut-être un véritable système dans lequel le guidage

prédéterminé de type mécanique est remplacé par une dynamique de champ globale[64]. »

Cette citation fait écho à la critique sévère de la cybernétique que Paul Weiss vient de présenter au symposium Alpbach de 1968, « Beyond Reductionism ». Le contexte idéologique est alors celui d'une glorification de la théorie générale des systèmes. Son fondateur, le biologiste Ludwig von Bertalanffy, est présent. Dans un style qui eût été inconcevable aux conférences Macy, celui-ci fait le procès de la science « mécaniste » et « réductionniste », en y incluant le behaviorisme, la psychanalyse, la biologie moléculaire, la conception du cerveau comme ordinateur, toutes approches « physicalistes »jugées coupables de ne jamais poser les problèmes de la complexité organisée, des processus interactifs, des comportement finalisés et de l'information vue comme néguentropie. A l'écouter, on brûle de savoir dans quel camp il range la cybernétique, science à la fois « mécaniste », « réductionniste » et « physicaliste », et dont l'objet, précisément, inclut les comportements finalisés et l'information néguentropique ! On le découvre bientôt : « La plus célèbre des diverses approches systémiques est la cybernétique », dit-il, tout en regrettant que l'on réduise trop souvent à celle-ci la théorie générale des systèmes[65]. Dès 1968, à la veille de la mort de McCulloch, la cybernétique était donc déjà l'objet de toutes les confusions.

Cette identification de la cybernétique à une science systémique de la complexité est d'autant plus étonnante, de la part de Bertalanffy, qu'en 1955, discutant l'apport de la cybernétique à la question des totalités organisées, il avait, dans une veine très proche de l'embryologie, distingué entre les régulations primaires et secondaires dans les systèmes, reléguant leur aspect mécanique à une place subordonnée. Ce sont les régulations primaires qui sont les plus fondamentales, et elles se caractérisent comme « interaction dynamique » ; les régulations secondaires, qui prennent la forme de mécanismes de type feedback, leur sont simplement superposées. « Cet état de choses, écrit Bertalanffy, résulte d'un principe général d'organisation que l'on peut appeler la mécanisation progressive. Dans une première étape, les systèmes – qu'ils soient biologiques, neurologiques, psychologiques ou sociaux – sont gouvernés par l'interaction dynamique de leurs composants ; plus tard se mettent en place des arrangements fixes et des contraintes qui rendent le système et ses parties plus efficaces, mais qui diminuent aussi graduellement, et pour finir abolissent, son équipotentialité. Ainsi, c'est la dynamique qui constitue l'aspect le plus fondamental, puisqu'il est possible d'arriver à une fonctionnalité de type machinique en partant des lois générales des systèmes et en introduisant des contraintes appropriées, alors que l'inverse n'est pas possible[66]. »

C'est donc dans ce cadre « systémique » que Weiss précise, lors du symposium Alpbach, sa conception des systèmes auto-organisés. Parlant des processus de la synthèse cellulaire, il utilise des expressions comme systèmes *self-contained*, *self-perpetuating*, *self-sustaining*. Il s'agit de condamner l'approche en termes de *programme* génétique, en soulignant que ces processus exigent, pour pouvoir se dérouler, de disposer de leurs propres productions finales[67] . Contrairement aux cybernéticiens, Weiss établit une coupure très nette entre les « machines » et les « systèmes ». Ces derniers sont des totalités naturelles et non nominales : « Les systèmes sont les produits de notre expérience avec la nature, et non pas des constructions intellectuelles. » Et cependant, ce ne sont pas des monades, et leur reconnaissance ne relève pas du « holisme », au sens où celui-ci ferait des touts des substances ou des essences. Nous ne sommes donc pas revenus à Aristote ou à Leibniz et à leur vieil organicisme. Weiss explique pourquoi il préfère ne

pas suivre Gerard et Koestler lorsqu'ils caractérisent la « systémicité » des systèmes par des substantifs (respectivement : « orgs » et « holons »). Ces termes pourraient accréditer l'idée que les touts sont des substances détachables de la dynamique totalisatrice[68] . C'est d'un nouvel organicisme qu'il s'agit, propre à des totalités qui ne sont ni des monades, ni de simples agrégats. Cette démarche qui n'est ni holiste ni réductionniste, mais vise à « réconcilier » le réductionnisme et le holisme en adaptant la science moderne à l'« organisation complexe »[69], non seulement, donc, ce n'est pas la cybernétique qui l'aura inventée, mais c'est contre elle que, d'abord, elle aura été formulée.

Weiss n'a pas de mots trop durs pour critiquer l'utilisation abusive des métaphores cybernétiques en biologie moléculaire. Parlant des biologistes qui interprètent l'ordre du vivant comme l'expression de l'information génétique, il écrit : « Au lieu d'inverser la rigoureuse méthodologie objective – qui a permis de descendre avec succès de l'organisme au gène – et de chercher dans quelle mesure on aurait pu effectivement synthétiser un système d'ordre supérieur quelconque en partant seulement des gènes interagissant librement dans un environnement dépourvu d'ordre, ils ont eu recours à de prétentieux termes anthropomorphes, qui ont tourné le problème et l'ont obscurci. On s'est contenté de douer les gènes de la faculté de spontanéité, du pouvoir de « dicter », d'« informer », de « réguler », de « contrôler », etc. les processus sans ordre d'un milieu non organisé, de manière à mouler celui-ci dans un travail d'équipe coordonné qui parvient finalement à un organisme fini[70]. »

Système et autonomie

Revenons à l'année 1948 et au symposium Hixon. Le terrain avait bien été déblayé par la critique purement neurophysiologique de Karl Lashley. Le signal entrant par les organes des sens, explique-t-il en substance, ne pénètre pas dans un système nerveux dont la majeure partie des neurones est inactive ou au repos, comme le veut la théorie neurologique dominante. Au contraire, il faut voir l'écorce cérébrale comme un immense réseau de circuits réverbérants constamment en activité. En l'absence de stimulation externe, ce réseau se stabilise en formant des *patterns* d'activité réguliers qui remplissent de très nombreuses fonctions d'intégration, motrices, temporelles, syntaxiques, etc. C'est sur ce substrat qu'agissent les impulsions afférentes, par une réorganisation des « patterns » préexistants, et non pas en suivant des chemins précâblés pour produire un *output*. L'*output* est le résultat de l'interaction entre un événement perturbant et une forme globale qui intégre tout un système de neurones interagissant de façon spontanée. Comme le dira Simondon, l'information produite « n'est pas de la forme, ni un ensemble de formes, elle est la variabilité des formes, l'apport d'une variation par rapport à une forme ». Mais ici, c'est à McCulloch que s'adresse la leçon[71] . A l'appui de ses affirmations, Lashley cite ses fameuses expériences sur le rat, à qui l'on peut enlever des parties importantes du cortex cérébral sans que son comportement et ses capacités en soient altérés profondément. Il en conclut que l'on ne peut, comme McCulloch, soutenir la thèse d'une spécialisation spatiale des fonctions et des habitudes[72].

Aujourd'hui, Jean-Pierre Changeux insiste à son tour sur cette activité spontanée. Il constate que l'« activité évoquée [par une stimulation périphérique] ne constitue en fait qu'une faible fraction de l'activité totale observée en l'absence de stimulation sensorielle

évidente ». Il écrit, ce qui consone tout à fait avec l'argumentation de Lashley : « La variation d'un paramètre physique de l'environnement se trouve donc traduite en une variation d'impulsions nerveuses [...]. Une chaîne de réactions successives, explicables en termes strictement physico-chimiques, assure, de la surface sensible à l'oscillateur, *le réglage d'une activité spontanée qui préexiste à toute interaction avec le monde extérieur*. Ces impulsions produites sont donc de nature indépendante du paramètre physique auquel l'organe est sensible. Les organes des sens se comportent comme des "commutateurs" d'horloges moléculaires. Les stimuli physiques qu'ils reçoivent du monde extérieur les avancent, les retardent ou les remettent à l'heure. Aucune "analogie" physique n'existe entre le paramètre physique reçu de l'environnement et le signal nerveux produit[73]. »

Il revint à Weiss, lors du symposium Hixon, de formuler une critique semblable à celle de Lashley dans un langage qui est celui de la théorie des systèmes auto-organisateurs. Aucune théorie du système nerveux ne peut prétendre rendre compte des faits, dit-il, si elle méconnaît sa caractéristique principale : *l'autonomie*. McCulloch, dans sa conférence, avait pris l'exemple de la pression que l'on exerce sur le globe oculaire et qui nous fait voir de la lumière alors qu'il n'y a pas de lumière. Les signaux propositionnels véhiculés par l'influx nerveux sont alors faux, avait-il affirmé. Weiss lui réplique qu'il n'y a pas plus de fausseté dans ce cas que dans celui d'une excitation visuelle externe. Dans les deux situations se manifeste la même capacité constitutive de l'appareil optique de produire une activité que nous percevons comme lumière, qu'elle corresponde ou non à une stimulation externe. Deux conceptions radicalement différentes s'opposent ici, précise-t-il : celle qui fait du système nerveux central une simple machine à transformer des messages entrants en messages sortants ; et celle qui y voit un système qui a sa cohérence interne propre, dont les caractéristiques ne sont pas le reflet des stimulations sensorielles. La question de savoir si cette autonomie d'activité est engendrée par des circuits neuronaux, comme le suppose Lashley, ou si elle résulte d'un processus métabolique qui s'exprime sous la forme d'un champ électrique global, comme le postule Gerard, est ici secondaire. Ce qui importe, c'est de porter enfin le « coup de grâce » à la vieille idée de la psychologie associationniste selon laquelle c'est la structure de l'*input* qui détermine la structure de l'*output*. Ce qui ne veut pas dire, bien sûr, que le système nerveux est coupé du monde extérieur. Simplement, l'*input* n'agit que comme détente ou déclencheur, qui « choisit » entre divers modes de fonctionnement autonome du système et éventuellement les altère[74]. On ne peut qu'admirer la clairvoyance de ces propos tenus en 1948, et qui annoncent ce que sera l'esprit des recherches connexionnistes de type ANN (*Attractor Neural Network*), dont j'ai montré, au chapitre précédent, combien elles se prêtent à une interprétation phénoménologique. Mais, je le répète, c'est contre le projet cybernétique qu'ils furent tenus.

Au symposium Hixon, McCulloch était donc apparu isolé, coupable d'« atomisme[75] » et de réductionnisme. Au symposiuim « Beyond Reductionism », Weiss devait revenir sur sa critique des explications atomistes, pour leur substituer le principe d'un « déterminisme stratifié ». Ce principe traite l'auto-organisation du vivant comme l'émergence spontanée d'un ordre collectif stable à partir d'une population d'entités en interactions variables et imprécises. Ainsi que l'écrit Weiss, « considérant la cellule comme une population d'éléments de grandeurs diverses, l'ordre qui y règne est objectivement décrit par le fait que le comportement résultant de la population dans son ensemble est infiniment moins

variable d'un instant à un autre que les activités de ses éléments[76] ».

Tout organisme vivant est un système hiérarchique, explique notre embryologiste, dont les niveaux d'intégration supérieurs ne sont pas réductibles aux niveaux inférieurs. Il y a des « discontinuités » entre les niveaux. L'anti-réductionnisme de Weiss ne se présente cependant pas comme un « holisme », qui serait tout aussi simpliste que l'atomisme dans la mesure où il substituerait un niveau ultime d'explication à un autre, remplaçant les éléments par le tout. Même s'il n'utilise pas l'expression, Weiss énonce un principe de *causalité circulaire* entre niveaux d'intégration emboîtés. Il ne peut évidemment pas l'utiliser, puisqu'elle symbolise pour lui la cybernétique, qui est précisément la cible de sa critique. Il est cependant indéniable, paradoxalement, qu'elle représente beaucoup mieux la pensée de Weiss que celle de la cybernétique. Dans un « système » (c'est-à-dire un organisme), les lois de la physique laissent aux éléments individuels de nombreux degrés de liberté. Cette indétermination à la base va être réduite par les contraintes exercées par le tout, lesquelles résultent elles-mêmes de la composition des activités élémentaires. Le tout et les éléments se déterminent mutuellement. C'est cette codétermination qui explique la complexité des êtres vivants[77]. Francisco Varela ne dira pas autre chose en formulant ses « principes de l'autonomie biologique ». Ce qu'il faut ici retenir, c'est que la cybernétique aura manqué cela, alors même qu'elle plaçait ses conférences fondatrices sous le signe de la « causalité circulaire ».

On est d'ailleurs effaré, à la lecture des Actes des conférences Macy, de constater combien le thème de la circularité est absent. Bien sûr, on parle de feedback, bien sûr on mentionne régulièrement l'existence et le rôle des circuits neuronaux « réverbérants », mais ce ne sont là que les formes les plus élémentaires de la causalité circulaire. Au-delà, on trouve bien peu de chose. Dès la première conférence, Wiener suggère d'injecter un paradoxe russellien dans un ordinateur, ce qui devrait le plonger dans un régime d'oscillations sans fin. Idée qui séduira beaucoup Bateson, puisque discutant à la neuvième conférence des mécanismes de l'humour, il avouera que cela ne le gêne en rien de dire d'un vibreur électrique qu'il rit[78] . Tout cela est bien innocent. L'index des cinq dernières conférences est à cet égard éloquent. A « circulaire », on ne trouve qu'une référence, à savoir la communication d'un célèbre critique littéraire de l'époque, Ivor Richards, qui parla à la huitième conférence du type de langage qu'il faut utiliser pour parler du langage. Discussion très verbeuse, mais qui contient la seule référence au théorème d'incomplétude de Gödel de tous les Actes publiés. Cette référence fut de pure forme, et sans aucune productivité[79] . On ne compte qu'une entrée en *self,* – qui renvoie à une remarque faite en passant par von Foerster sur la possibilité d'une *self-communication*[80] . Lorsque dans sa recension de l'ouvrage de Steve Heims, *John von Neumann and. Norbert Wiener,* von Foerster reproche à celui-ci de ne pas avoir redonné à la causalité circulaire toute la place qui fut la sienne dans l'histoire de la cybernétique, il semble bien qu'il confonde ce que nous appelons ici la première et la seconde cybernétiques[81].

Le brouillage de la notion de modèle, ou le virus de la complexité

Comme l'a bien vu Gilbert Simondon, les limites de la cybernétique ont surtout tenu à

ce qu'elle est restée jusqu'au bout prisonnière de son postulat initial : « l'identité des êtres vivants et des objets techniques auto-régulés[82] ». Ce n'est pas, contrairement à ce qu'on affirme parfois, que les cybernéticiens aient été principalement des techniciens ou des ingénieurs. S'ils s'intéressaient tant aux machines, ce n'était pas en tant qu'*applications* utiles d'un savoir scientifique déjà constitué, c'est qu'elles représentaient à leurs yeux l'incarnation dans la matière d'hypothèses ou de théories de type logico-mathématique.

Lors de la huitième conférence Macy, Shannon présenta son automate, qu'il avait baptisé « rat ». Le groupe cybernétique tomba sous le charme. Von Foerster rend compte de sa fascination en ces termes : elle « ne résulte pas d'une quelconque similitude entre la machine et un vrai rat [...]. Mais il est frappant d'observer la ressemblance entre le mécanisme et les *notions* que certains théoriciens de l'apprentissage développent au sujet des rats et des organismes en général[83] ».

La cybernétique est peut-être la « métaphysique de l'âge atomique », elle est en tout cas le point d'aboutissement du principe selon lequel *Verum et factum convertuntur*, comme nous avons essayé de le montrer au premier chapitre. Le témoignage de Jerome Lettvin au sujet de l'évolution intellectuelle de Warren McCulloch après 1952 nous en apporte une confirmation troublante. En un chassé-croisé curieux, alors que Pitts le logicien délaissait les modèles formels au profit de la recherche expérimentale sur le système nerveux, McCulloch le neurophysiologiste allait consacrer tous ses efforts à la théorie et à la logique. Déçu par les méthodes des neurosciences, nous dit Lettvin, « il se vouait à la tâche de savoir comment le cerveau fonctionne dans les termes mêmes où le créateur d'une machine connaît ses rouages. La clé d'un tel savoir n'est pas dans l'observation, mais dans la fabrication de modèles, que l'on confronte ensuite aux données. Mais la *poiesis* doit venir d'abord. Et McCulloch préférait risquer l'échec dans sa tentative de créer un cerveau que rencontrer la réussite dans l'amélioration de la description des cerveaux existants[84] ». Lettvin ajoute que ce n'est pas par hasard que McCulloch était alors entouré de ceux qui, tels Seymour Papert et Marvin Minsky, allaient fonder l'intelligence artificielle.

Cette évolution, que l'on peut décrire comme l'émancipation progressive par rapport au réel de l'outil par lequel le savant recrée le monde, était en un sens annoncée dans un texte de 1945 que Wiener et Rosenblueth publièrent dans la revue qui avait accueilli leur article fondateur de 1943. Intitulé « The Role of Models in Science[85] », ce texte commençait par rappeler l'importance qu'a toujours eue l'activité de modélisation dans la science moderne. A défaut de pouvoir se rendre maître et possesseur d'une portion du monde, le savant en construit une image plus simple et aussi fidèle que possible dont, comme d'un fétiche, il s'assure la maîtrise. Or, tout au long de l'article, les auteurs traitent non seulement des modèles formels, pures constructions intellectuelles, mais aussi des modèles matériels, qui sont des objets du monde, comme l'objet qu'ils représentent. Pour finir, ils considèrent le « cas limite » de la modélisation : un objet dont le seul modèle matériel satisfaisant serait cet objet lui-même. Il se trouve que cette figure, que Wiener a peut-être empruntée à Josiah Royce dont il suivait le séminaire à Harvard dans les années 1911-1913[86] , va hanter les conférences Macy, parfois de façon très explicite. En témoigne l'échange entre Ralph Gerard et McCulloch à la neuvième conférence. Ironie ou fausse modestie, le premier explique que ce que le groupe, au fond, semble attendre des neurophysiologistes, c'est qu'ils fournissent aux mathématiciens des données pour que ceux-ci puissent bâtir leurs modèles sans perdre trop le sens du réel. Le réel est

devenu le moyen, et le modèle la fin. A quoi McCulloch réplique : « Le meilleur modèle du comportement du cerveau, c'est le comportement du cerveau lui-même[87]. » Mais c'est le plus souvent de façon non réfléchie que l'objet et le modèle éprouvent une fâcheuse tendance à constamment inverser leurs rôles. Oscillation que le mot même de modèle, comme nous l'avons dit, recèle dans son ambivalence : est modèle ce qui imite, mais aussi ce qui est, ou mérite d'être imité.

Typique à cet égard est, dans son extrême confusion qui mit certains participants hors d'eux, la discussion de la septième conférence sur le numérique et l'analogique. Nous l'avons déjà évoquée comme l'un des moments chauds où s'affrontèrent les partisans d'une conception « continuiste » du système nerveux et leurs adversaires « neuronistes ». A ce niveau, analogique veut dire continu et numérique, discret. Plus précisément, comme l'expliquent von Neumann et Pitts, si l'action d'une variable continue sur la dynamique d'un système dépend seulement du fait que sa valeur dépasse ou non un certain seuil, alors on peut *décrire* le rôle qu'elle joue dans l'économie du système comme étant celui d'une variable « numérique », du type oui ou non, ou 0 ou 1 ; sinon, cette variable sera dite « analogique ». Dans ce premier sens, on peut donc dire d'une réalité naturelle qu'elle est analogique. Cependant, puisque l'on parle aussi et surtout de machines, que l'on conçoit comme des modèles, des représentations d'une réalité naturelle, il est inévitable que le mot « analogique » évoque non plus un rapport interne au système de la machine, mais un rapport entre celle-ci et l'objet naturel qu'elle est censée représenter. Rien n'interdit dès lors à une machine d'être analogique dans ce second sens et numérique dans le premier. Mais il s'agit du système nerveux, qui est un objet naturel bien particulier puisqu'il est l'organe même de toute représentation. A son sujet, les deux sens viennent en coalescence, et l'on ne sait bientôt plus ce que l'on dit lorsque l'on traite des représentations artificielles, matérialisées, du cerveau[88] . C'est la notion même de modèle qui, dans la discussion, s'estompe progressivement, dans la mesure où elle suppose une relation hiérarchique entre une origine et une reproduction. Entre les faits et leurs descriptions, il n'y a bientôt plus de différence, la hiérarchie fait place à une relation d'équivalence, qui est l'adéquation à un même modèle formel[89] . C'est la logique qui met à plat le monde et ses représentations. Les modèles de la cybernétique sont déjà poststructuralistes, ils ne sont modèles que d'eux-mêmes, ou bien d'autres modèles, miroirs de miroirs, spéculums ne réfléchissant aucune réalité[90].

Le seul à avoir maîtrisé cette dérive, c'est-à-dire à avoir tenté de lui trouver une justification rationnelle, est von Neumann. Nous savons par Burks qu'il vint à s'intéresser au calcul parce qu'il butait sur les problèmes de la dynamique non linéaire, en particulier dans ses recherches sur la dynamique des fluides. Il pensait, par exemple, que la chimie ne pourrait accéder au rang de science pure que le jour où l'on saurait intégrer les équations de la mécanique quantique. En attendant, presque tout ce que nous savons au sujet des solutions de ces équations, disait-il, nous le devons aux expériences que nous faisons sur les systèmes physiques eux-mêmes. Singulier renversement, puisque, si l'on admet par définition que le modèle est plus simple et mieux maîtrisable que l'objet qu'il représente, alors c'est ici l'objet naturel qui sert de modèle à l'objet mathématique, et non l'inverse.

Ce basculement de l'idée de modèle est contemporain du mouvement par lequel la science se concentre sur l'étude des objets « complexes », au sens que von Neumann a donné à ce mot. Le modèle mathématique, artificiel, perd son statut d'instrument de

maîtrise, puisqu'on ne sait pas le *résoudre*. La complexité de l'objet réel est irréductible, et seule l'histoire de cet objet dans le monde peut nous dire ce dont il est capable. Cette idée révolutionnaire, que les sciences de la nature (et de l'artificiel) découvrent à cette époque, est en réalité une vieille connaissance de la pensée du social. On ne devrait peut-être pas s'en étonner, tant il est vrai qu'aucun objet du monde physique ou mécanique n'atteint la complexité d'une société. Ce n'est pas un hasard – nous y reviendrons – si un Friedrich von Hayek se réfère à la fin des années trente au passage suivant du *Manuel d'économie politique* de Vilfredo Pareto. Celui-ci écrit en 1906 que le modèle de l'équilibre économique général, mis au point par Léon Walras et par lui-même, et qui formalise le mécanisme de la formation des prix sur un marché concurrentiel, « ne vise nullement à un calcul numérique des prix. Faisons en effet l'hypothèse la plus favorable pour un tel calcul, supposons que nous ayons triomphé de toutes les difficultés consistant à trouver les données du problème et que nous connaissions les *ophélimités* [c'est-à-dire les « utilités » ou « désirabilités », JPD] de tous les différents produits pour chaque individu, et toutes les conditions de production de tous les produits, etc. Cela est déjà une hypothèse absurde. Pourtant elle n'est pas suffisante pour rendre la solution du problème possible. Nous avons vu que dans le cas de 100 personnes et de 700 produits il y aura 70 699 conditions (en fait, un grand nombre de circonstances que nous avons négligées jusqu'ici augmenteront ce nombre) ; nous devrons donc résoudre un système de 70 699 équations. Cela excède en pratique les pouvoirs de l'analyse algébrique, et cela est plus vrai encore si l'on songe au nombre fabuleux d'équations que l'on obtient pour une population de quarante millions d'habitants et plusieurs milliers de produits. *Dans ce cas, les rôles s'inverseraient* : ce ne seraient plus les mathématiques qui viendraient au secours de l'économie politique, mais l'économie politique qui viendrait au secours des mathématiques. En d'autres termes, si l'on pouvait réellement connaître toutes ces équations, le seul moyen pour les résoudre qui soit accessible aux pouvoirs de l'homme serait d'observer la solution pratique qui leur est donnée par le marché[91] ».

Le contexte idéologique, en tout cas pour ce qui est de Hayek, est clair : seul le marché peut nous dire ce dont il est capable. Le meilleur modèle, et le plus simple, du comportement du marché, c'est le comportement du marché lui-même. L'information que le marché mobilise et met au service de ceux qui se laissent porter par sa dynamique, n'est pas « compressible ». Nul, donc, ne peut s'en dire le détenteur, ni prétendre commander aux « forces aveugles du processus social » – aveugles, mais bénéfiques. La critique hayékienne du « rationalisme constructiviste » en philosophie sociale anticipe la critique adressée par von Neumann à la philosophie artificialiste de McCulloch.

Le modèle mathématique qui décrit un objet complexe est lui-même un objet complexe. Un de ses « modèles » possibles, nous venons de le voir, c'est l'objet réel lui-même. Mais grâce à von Neumann, l'apparition de l'ordinateur ouvre une nouvelle possibilité : la *simulation* numérique. A son tour, l'objet mathématique est traité comme un objet naturel. Sa définition n'est pas « généti que », c'est-à-dire que l'on ne peut déduire de la définition toutes les propriétés de l'objet. Pour connaître celui-ci, il est donc indispensable de recourir à l'expérimentation, sur la représentation matérialisée dans l'ordinateur. Très tôt, von Neumann fut donc sensible à ridée que nos constructions intellectuelles, dès lors qu'elles sont « complexes », ne sont pas d'une nature radicalement différente de celle des objets du monde physique. Cela ne pouvait que le préparer à « déconstruire » la relation hiérarchique inhérente à la notion de modèle. C'est en tout cas

cette position philosophique qui fournit à von Neumann la motivation et l'énergie nécessaires à la conception et au développement des calculateurs électroniques[92].

C'est dans ses travaux sur la théorie des automates que von Neumann devait préciser cette notion de complexité. En faisant une grandeur de type thermodynamique, il conjecturait qu'en deçà d'un certain seuil, elle est dégénérative, le degré d'organisation ne pouvant que décroître, mais qu'au-delà la *complexification* n'est pas une notion contradictoire. Or ce seuil de complexité, supposait-il, est aussi celui où la structure de l'objet devient plus simple que la description de ses propriétés. Dans le cas habituel, celui des machines simples, il est moins compliqué de dire avec des mots ce dont l'automate est capable que de présenter son plan de câblage. Pour des automates complexes, l'inverse serait vrai : il serait plus simple, voire infiniment plus simple, de décrire la structure de l'automate que de décrire complètement son comportement. C'est sur cet argument que von Neumann fera reposer sa critique du modèle de McCulloch et Pitts au symposium Hixon, comme nous l'avons vu. A l'appui de sa conjecture, il invoquait certains résultats de la logique formelle et avouait qu'il faisait subir aux idées de Gödel une légère altération : « C'est un théorème de Gödel que l'étape logique suivante, la description d'un objet, est d'un type de classe plus élevé que l'objet et est donc asymptotiquement infiniment plus longue à décrire[93]. » Ce texte, édité par Burks à partir d'une conférence donnée en 1949, n'est pas d'une clarté admirable. Par « description d'un objet », il faudrait comprendre « description du comportement de l'objet », et par « objet », « structure de l'objet ». On peut interpréter la négligence manifestée ici par le grand mathématicien comme un signe du brouillage qu'il fait subir à la notion de modèle.

Gödel, consulté par Burks après la mort de von Neumann, proposera de justifier l'assertion de son collègue de Princeton de diverses façons (tout en reconnaissant n'avoir jamais discuté de ce sujet avec lui !)[94] . Il songe d'abord au théorème de Tarski sur la vérité de 1931. Dans tout langage consistant, il est impossible de caractériser l'ensemble des phrases vraies de ce langage par un prédicat purement syntaxique. La sémantique est donc irréductible à la syntaxe. C'est là, précise Gödel, la « vraie raison » du théorème d'incomplétude. Il est significatif que, bien plus tard, en 1970, Donald Davidson, posant les prolégomènes à son monisme non réductionniste ou monisme anomal, ait eu recours à ce même théorème de Tarski pour illustrer, à titre de métaphore, ce qu'il tient pour être le dualisme irréductible des propriétés mentales et des propriétés physiques : le mental est au physique ce que la sémantique est à la syntaxe[95].

La conjecture de von Neumann sur la complexité, elle aussi, ouvre la voie à un matérialisme non réductionniste, certes fort différent de celui de Davidson. Elle rend, par exemple, non contradictoires les deux propositions suivantes : 1) Des mécanismes physico-chimiques sont capables de produire la vie ; 2) La vie est (infiniment) plus complexe que les mécanismes physico-chimiques qui l'ont engendrée. Il est cohérent d'embrasser une ontologie tout à la fois non substantialiste (ici, non vitaliste) *et* non réductionniste. Cette conclusion est remarquable si l'on note que les ontologies non réductionnistes sont très généralement substantialistes, et les ontologies non substantialistes presque toujours réductionnistes.

Cependant, note Gödel, il est plus immédiat de justifier la conjecture de von Neumann par les propriétés de la machine de Turing universelle. Du fait de l'indécidabilité de l'arrêt, il n'y a pas de moyen plus économe de décrire son comportement que d'énumérer toutes ses productions possibles (une infinité). De tous les cybernéticiens, von Neumann

aura été le seul à être sensible à cet aspect du travail de Turing, et de la logique formelle en général. Alors que les autres invoquaient à tout bout de champ la thèse de Turing (trop vite baptisée « théorème »), comme pour affermir leur foi dans la réductibilité du comportement humain au fonctionnement d'une machine, lui prenait la perspective inverse, et retenait la leçon proprement paradoxale que le comportement d'une machine n'est pas la chose banale que l'on croit puisque aucune machine ne peut le prévoir complètement.

Il n'est pas difficile d'imaginer les conclusions que von Neumann pouvait tirer de ses idées sur la complexité en ce qui concerne la philosophie de la modélisation. Au symposium Hixon, alors que les neurophysiologistes (y compris McCulloch) lui reprochent de ne pas assez faire la différence entre les automates naturels et les automates artificiels, il réplique que cette distinction ira de plus en plus en s'affaiblissant. Bientôt, prophétise-t-il, le constructeur d'automate sera aussi désarmé devant sa création que nous le sommes devant les phénomènes naturels complexes[96]. Le modèle, qui était hiérarchiquement subordonné au réel qu'il ne faisait que mimer, s'émancipe et devient l'égal de son référent. Soit un objet naturel que l'on cherche à modéliser, et dont la complexité est au-delà du seuil critique postulé par von Neumann. Le modèle, s'il doit être fidèle à ce qu'il représente, devra lui aussi dépasser ce seuil. Mais alors, il sera non seulement modèle de son objet, mais aussi modèle de lui-même, ou plutôt de son comportement. Il n'en faut pas davantage pour que le modèle ne renvoie plus qu'à lui-même, polarisant sur lui l'attention, au détriment de l'objet originel. C'est ainsi que la neurophysiologie laisse la place à l'intelligence artificielle. L'autonomisation du modèle se confond avec la modélisation de l'autonomie. Griserie et paradoxe de fabriquer un automate, c'est-à-dire un être qui ne tienne sa loi que de lui-même. Comble de la volonté de puissance que de se vouloir la cause d'un être qui soit cause de soi, inconditionnée ; et, simultanément, perte de soi dans le miroir tendu par une créature faite à son image.

6

Aspects d'une déconvenue

« Le monde symbolique, c'est le monde de la machine. »

Jacques LACAN (1954)

Pour celui qui écrit, avec une attitude faite *a priori* de sympathie et même de gratitude – comme c'est le cas de l'auteur de ces lignes –, l'histoire de la cybernétique, le sentiment qui finalement prévaut est celui d'une grande déception. La nouvelle discipline se voulait l'avant-garde de la démarche scientifique, tant par son objet – l'esprit, ce chef-d'œuvre de la création – que par ses concepts, ignorés jusque là de la physique – la téléologie, l'information, la causalité circulaire, le feedback, etc. – et son style – réflexif, c'est-à-dire réfléchissant l'emploi de ses outils conceptuels. Est-ce l'arrogance liée à la certitude de faire franchir à la science une étape décisive, ou simplement le type de fermeture d'esprit que tout « paradigme » scientifique, nouveau ou ancien, requiert de ses membres ? Toujours est-il que la cybernétique aura – quel gâchis ! – raté l'essentiel de ses rendez-vous. J'ai décrit et analysé l'échec de ses rencontres avec la biologie, qu'il s'agisse de la biologie moléculaire, de l'embryologie ou de la neurophysiologie ; avec la psychologie, traitée au mieux comme un territoire à conquérir ; avec la phénoménologie, dont la « philosophie de l'esprit » eût mieux convenu à son modèle que cette branche de la philosophie analytique qui porte aujourd'hui ce nom. On peut aussi mentionner l'absence de contact avec la physique des systèmes désordonnés qui pourtant, mais bien plus tard, allait contribuer à fonder le *néo*-connexionnisme. Je voudrais conclure cette étude par l'analyse brève de deux échecs supplémentaires et, historiquement, très importants : l'échec à se saisir des perspectives ouvertes par la notion de complexité ; la fermeture à peu près totale aux ressources offertes par les sciences de l'homme et du social.

Nul ne peut cependant nier le lien de parenté entre le projet cybernétique et les nombreuses et révolutionnaires entreprises scientifiques et philosophiques qui lui succéderont, parfois en se recommandant de son nom, souvent en préférant le taire, de la théorie générale des systèmes aux sciences cognitives en passant par la biologie moléculaire. Que ces entreprises aient parfois affirmé leur identité *contre* la cybernétique ne saurait faire oublier qu'elles doivent en définitive la vie à l'initiative que prirent les hommes et les femmes qui constituèrent le groupe cybernétique de se réunir régulièrement, à l'hôtel Beekman, Park Avenue, à New York City, pour tenter de bâtir ensemble une nouvelle science de l'esprit.

Une dernière note de tristesse. Ceux qui, aujourd'hui encore, à la fin du XXe siècle, se reconnaissent dans l'étendard cybernétique, se réunissent le plus souvent à l'invitation de groupes de psychothérapie familiale. C'est certes une discipline respectable, mais que ce soit pratiquement la seule désormais à abriter les rencontres cybernétiques fait désagréablement écho au malentendu qui présida aux conférences Macy. De la paix du

monde on est simplement passé à la paix des ménages. Les femmes et les hommes qui, les premiers, crurent possible d'ouvrir à la démarche scientifique le domaine de l'esprit, méritaient mieux.

L'apprentissage de la complexité

Von Neumann avait posé le problème de la complexité, mais ce n'est pas sous son influence que les cybernéticiens s'y trouvèrent confrontés. Certes, nous l'avons dit, McCulloch en fut ébranlé, d'autant que dès 1946, avec Pitts, il avait en quelque sorte inversé de lui-même les termes de son propre problème. Dans sa conférence « Finality and Form in Nervous Activity », il rappelait leur résultat de 1943, à savoir que toute « idée » est incarnable dans un réseau de neurones. Mais de tels réseaux, construits très précisément pour accomplir telle ou telle performance, existent-ils réellement dans le cerveau ? demandait-il, pour répondre que c'était peu probable, car les gènes n'ont pas la capacité suffisante pour spécifier le plan de câblage du système tout entier. Il faut donc partir de réseaux *aléatoires*, concluait-il, et s'interroger sur le type de propriétés qu'ils sont capables d'engendrer[1] . Il n'est cependant pas certain que ce changement méthodologique, pour remarquable qu'il fût, se soit accompagné d'un changement équivalent dans le rapport à l'objet scientifique. On le voit bien lorsque, en réponse à Ashby, à la neuvième conférence, Pitts conjecture que la thèse de Turing reste vraie dans le cas aléatoire[2]. On n'est donc pas sorti du cadre de pensée « artificialiste » ou « constructiviste », qui consiste à se donner une fonction et à en *déduire* la structure capable de la réaliser.

McCulloch était le premier à revendiquer cette démarche « descendante » (*top-down*, comme on dit en anglais) qui va des propriétés du tout à celles des éléments et de leurs connexions. La question lui fut posée, au symposium Hixon, de savoir quel était son intérêt premier, du système nerveux ou des machines à calculer. Sa réponse : il n'y a pas entre eux de différence de conception (*design*), dès lors que les tâches qu'ils sont capables d'exécuter *déterminent* leur structure[3] . Or, simultanément, on lui reprochait son « atomisme » – cette démarche ascendante (*bottom-up*), qui consiste à partir d'éléments complètement spécifiés ainsi que de leurs connexions, et à en *déduire* les propriétés de la totalité qu'ils constituent – et lui-même se reconnaissait dans cette présentation. Cela apparaît nettement lorsque, par exemple, il affirme que l'essentiel de la *Gestalt*-psychologie est « déductible » du type de neurophysiologie qu'il pratique[4].

Loin d'être contradictoire, cette réversibilité du mouvement de la pensée entre le tout et ses parties est la marque de ce que l'on peut appeler un abord « simple » du réel. Si la totalité se déduit simplement des éléments, c'est que ceux-ci contiennent déjà l'intelligence de la totalité. Ils sont intelligents, trop intelligents, puisqu'ils ont été conçus *pour* la totalité. On le verra bien avec la biologie moléculaire et ses métaphores cybernétiques : les gènes et les enzymes y seront dotés de toutes les qualités de la pensée et de la volonté humaines, puisqu'ils seront capables d'informer, de réguler, de commander, de transcrire et de traduire[5]. Par contraste, on peut dire que la conception de l'organisation vivante comme auto-organisation n'est ni de type « descendant », ni de type « ascendant », puisqu'elle implique codétermination du tout et des éléments. Le problème tel que Paul Weiss le posait et tel qu'il est repris aujourd'hui par les

néoconnexionnistes est celui des comportements collectifs stables dont est capable un réseau d'éléments interconnectés que l'on n'a défini que de façon « aléatoire ». Il est essentiel de comprendre qu'« aléatoire » ne veut plus dire ici que le système n'est pas déterministe, mais que l'on s'est donné au hasard les connexions. Elles ne sont donc pas destinées *a priori* à réaliser un projet collectif, c'est la totalité particulière qu'elles produiront effectivement dans telles ou telles circonstances qui, *a posteriori*, leur donnera sens.

Il se trouve que les cybernéticiens durent se mesurer à cette problématique si différente de la leur. Ce n'est pas qu'elle leur fut exposée de façon didactique par un théoricien. Ils la rencontrèrent dans la *pratique* de certains des scientifiques qu'ils invitèrent, des expérimentateurs en situation d'interaction avec la nature ou le vivant. Nous nous limiterons ici à revenir sur deux cas que nous avons déjà brièvement analysés dans une optique différente.

Le point de vue cybernéticien sur les expériences d'Alex Bavelas, on s'en souvient, se concentre sur la tâche à accomplir, et juge des performances réalisées par le groupe en fonction de cet objectif, les comparant à celles d'une machine. Qu'une information supplémentaire ait apparemment pour effet de rendre plus difficile la résolution du problème ne peut être jugé que comme un signe d'irrationalité. Mais que fait Bavelas, lui, en situation d'expérimentation ? Oubliant l'objectif qu'il a lui-même assigné à ses cobayes, il s'intéresse à la façon dont le groupe se structure, il explore les multiples formes qu'il peut prendre. La question intéressante n'est plus : comment le groupe va-t-il accomplir cette tâche que je lui impose ? mais : de quels comportements collectifs imprévus le groupe est-il capable ? Il n'y a plus dès lors à s'étonner qu'une information nouvelle accroisse la complexité des réponses possibles et, donc, le temps mis à résoudre le problème initial – mais ce n'est plus celui-ci qui est jugé *intéressant*.

Plus que toute autre, la discipline écologique aura été sensible et attentive à la question de la complexité. Theodore Schneirla, le spécialiste des colonies de fourmis, à la deuxième conférence, et Herbert Birch, à la huitième, en administrent la preuve aux cybernéticiens. Nous avons déjà fait écho à la démonstration de ce dernier et aux réactions violentes qu'elle suscita. Les comportements collectifs les plus complexes que l'on rencontre dans les sociétés animales, expliqua-t-il, qu'il s'agisse des abeilles et de la façon dont elles s'orientent ou communiquent avec leurs congénères, des fourmis et de leurs expéditions ou de leurs bivouacs, des poissons et de leurs bancs – tous ces comportements résultent en dernière analyse d'un asservissement à une certaine organisation sensorielle qui, si les circonstances changent, peut de la même façon produire les comportements les plus aberrants et les plus destructeurs. De l'« intelligence » apparente de l'ensemble, concluait-il, on ne saurait inférer l'intelligence des éléments : ceux-ci n'ont pas le comportement collectif pour projet[6] . Rappelons que si les cybernéticiens protestèrent, ce n'est pas parce que Birch traita ses animaux de stupides – mais parce qu'il tint à réserver le qualificatif d'intelligent à l'homme. On peut donc penser que les cybernéticiens commençaient à faire l'apprentissage de la complexité. Le fait qu'ils noyaient de plus en plus la notion de feedback dans celle de causalité circulaire en porte également témoignage[7]. Le *feedback*, en effet, incarne l'intelligence de la totalité, ce qui n'est pas nécessairement le cas d'une boucle dans un système complexe.

Le « cas Ashby », ou le retour à la métaphysique

Les cybernéticiens se familiarisaient ainsi, peut-être, avec la *pratique* de la complexité. Mais vint l'invité de la pénultième conférence, Ross Ashby. De l'œuvre de ce dernier, nous dirons peu de choses ici, parce qu'elle constitue l'une des clés de voûte de ce qu'on appelle la seconde cybernétique et qu'elle échappe donc au domaine principal de notre enquête. Nous nous contenterons de décrire ce que fut l'impact de son intervention sur les cybernéticiens première manière. Ceux-ci la reçurent comme un électrochoc, et leurs réactions en la circonstance ajoutent une touche intéressante, bien qu'ultime, au portrait que nous brossons d'eux.

Ashby, en un sens, aura pris tellement au sérieux le problème de la complexité que toute sa vie aura été consacrée à en extraire la quintessence. Ne cachant pas sa fascination pour Bourbaki, c'est dans une axiomatisation de la question qu'il en aura cherché la clé. Tout nous porte à penser que son influence a été, directement ou indirectement, considérable sur ceux de nos contemporains, et ils sont nombreux, qui ont rencontré sur leur chemin les idées de la cybernétique, qu'ils les aient ou non identifiées pour telles[8]. Le problème d'Ashby, c'est qu'il aura malgré lui administré la preuve qu'axiomatiser la complexité, c'est la faire disparaître en fumée.

A la neuvième conférence Macy, la seule à laquelle il prit part, Ashby présente deux communications : la première, sur l'automate qui allait le rendre célèbre, l'« homéostat » ; la seconde, sur le problème suivant : un joueur d'échecs mécanique peut-il être plus fort que son concepteur ? Oui, répond-il, il lui suffit de jouer au hasard. Soit en effet une configuration des pièces sur l'échiquier apparemment anodine, mais qui cache la possibilité d'un coup décisif. Deux classes de joueurs seulement pourront le jouer. Le débutant, d'abord, parce qu'il fait n'importe quoi ; le joueur aléatoire, ensuite, parce qu'aucune possibilité ne lui est *a priori* fermée.

C'est ce principe, si l'on ose dire, que l'homéostat incarne dans la matière d'une machine. Cet automate cybernétique a la prétention de représenter l'interaction d'un organisme – le « cerveau » – et de son environnement. Il ne sert aucun projet, n'est conçu pour remplir aucune tâche. C'est une « expérience de pensée » destinée à illustrer une thèse qu'Ashby voudrait universelle : loin d'être des accidents extrêmement improbables de l'évolution, la vie et l'intelligence se développent nécessairement dans tout système isolé. Être vivant, c'est avoir la capacité de maintenir un petit nombre de variables « essentielles » à l'intérieur de certaines limites physiologiques, et ceci pour une gamme très large d'environnements différents. Dans ce but, l'homéostat agit à deux niveaux. Au premier, toutes les connexions sont fixées, le système global est déterministe et sa dynamique figure la coévolution du « cerveau » et de son environnement. Si, au cours de cette évolution, une ou plusieurs variables essentielles du cerveau sortent de leur zone vitale, le second niveau est enclenché automatiquement, et l'homéostat change ses connexions *au hasard*. Il devient donc un *nouvel* automate déterministe, qui aura ou non la capacité de maintenir les variables essentielles dans la zone voulue. L'homéostat aura démontré la thèse qu'il incarne, s'il réussit, quel que soit l'environnement, à trouver les bonnes connexions, celles qui maintiennent le cerveau « en vie » et l'« adaptent », donc, à cet environnement.

Il y a un « truc » mais les participants à la conférence Macy ne le trouvent pas. Ce n'est pas qu'il est difficile, mais les auditeurs d'Ashby sont trop interloqués pour le voir. Ashby lui-même ne l'explicitera que plus tard, en l'axiomatisant[9]. Nous ne livrerons ici que l'indication suivante, juste destinée à illustrer sa méthode et son style. On le conçoit

aisément, l'« adaptation » exige que les variables essentielles n'occupent qu'un très petit nombre des états qui leur sont en principe accessibles. Or qu'est-ce qu'un automate ? Formellement, c'est une application mathématique d'un ensemble fini dans lui-même, que l'on réitère un nombre indéfini de fois. Or il y a peu d'applications « bijectives », c'est-à-dire qui conservent la cardinalité de leur ensemble de départ ; il y en a d'ailleurs proportionnellement d'autant moins que cette cardinalité est élevée. Cela veut dire que pour la très grande majorité des automates, le nombre d'états qui leur sont effectivement accessibles décroît très vite au cours de leur évolution. C'est là une conséquence triviale de la définition mathématique des applications (ou encore des fonctions), qui peuvent envoyer des sources différentes sur une même cible, mais, inversement, n'affectent jamais plusieurs cibles à une même source. Ashby aura donc finalement érigé une définition mathématique et son relatif arbitraire en principe universel de la nature.

Faute de voir cela, ses auditeurs l'attaquent sur ce qui leur semble le plus énorme : le recours au hasard pour simuler la pensée, l'invention, le génie, l'adaptation aux situations imprévues. Vous n'allez pas jusqu'à affirmer que le mouvement brownien est la meilleure façon pour un *organisme* d'inventer de nouvelles solutions ? lui demandent Quastler, Pitts et Bigelow. Puisque c'est une machine, je ne connais pas d'autre moyen, réplique Ashby. Mais enfin, insiste Bigelow, comment pouvez-vous affirmer que votre homéostat *apprend*, alors qu'il ne fait que *trouver* à tâtons une position d'équilibre ? Diriez-vous d'un roulement à billes qui finit par glisser à travers l'ouverture d'une boîte que l'on secoue à cet effet qu'il a *appris* à trouver la sortie ? Cela ne me gêne pas, répond Ashby, parce que apprendre, cela n'a pour moi qu'un sens objectif, qui n'a rien à voir avec l'introspection. Bigelow, alors, explose : le rat de Shannon, lui, est capable d'apprendre, mais pas votre machin[10].

Les rôles, on l'aura remarqué, sont parfaitement inversés. Ashby réplique aux cybernéticiens avec les arguments mêmes dont ceux-ci usaient à l'égard d'un Kubie ou d'un Birch. Ce sont eux qui, maintenant, en sont misérablement réduits à distribuer les bons et les mauvais points mentalistes. Ashby tend aux cybernéticiens un miroir déformant dans lequel ils aperçoivent une caricature d'eux-mêmes, et ils ne s'y reconnaissent pas. Ils ne font pas de coupure entre l'univers des machines et celui des organismes ? Ashby leur apprend que l'essence de la vie n'a rien qui lui soit propre[11] et qui ne puisse être incarné dans un dispositif technique composé de quelques électro-aimants interconnectés. Ce sont des champions de l'objectivité et de l'anti-mentalisme ? Chez Ashby, les mathématiques ensemblistes élémentaires prennent la place de l'objet et le hasard celle de l'algorithme. Ils abstraient de l'organisme la machine logique sous-jacente et leur démarche oscille entre la logique et la physiologie ? Ashby fait fond sur la notion de fonction et perd, et la logique, et la biologie. Ils apprennent à poser les problèmes de la complexité et de la complexification ? Ashby les incarne dans des « expériences de pensée » élémentaires et autocontradictoires, puisque leur simplicité « pédagogique » une fois découverte, il ne reste rien du bon tour qu'elles nous ont joué.

C'est avec Ashby que le principe du *verum factum* atteint véritablement son point d'aboutissement dans l'histoire de la science. Des concepts comme « cerveau », « esprit », « vie », « organisme », « évolution », « intelligence » perdent tout contact et tout rapport avec un quelconque donné, ils sont entièrement reconstruits dans le cadre d'un modèle mathématique axiomatisé. La forme s'arrache totalement et définitivement à la matière. McCulloch cherchait les conditions formelles *et matérielles* du jugement

synthétique *a priori*. Il se posait les questions que se posent les biologistes et les psychologues, même s'il avait recours à l'outil logico-mathématique. La même chose peut être dite au sujet de Herbert Simon et de ses systèmes symboliques *physiques*, c'est-à-dire matériels. Avec Ashby, on est dans le pur *a priori* mathématique. Le projet interdisciplinaire de la cybernétique, l'interaction avec ceux qui, concrètement, interrogent le vivant ou l'esprit et entrent en « conversation » avec les objets du monde, de difficile et limité qu'il était, est maintenant devenu radicalement impossible. Comme l'écrit Isabelle Stengers : « Il n'est plus question de dialogue [...]. La force d'Ashby est celle de la stratégie de la terre brûlée : là où il passe, il sera impossible de faire certaines hypothèses, ceux qui suivront devront se conformer aux contraintes qu'Ashby a énoncées pour l'éternité[12]. »

Même si elle n'en avait pas toujours conscience, la première cybernétique participait de la déconstruction de la vision métaphysique du monde. Avec Ashby, on y replonge à pieds joints. Le seul point de vue légitime est celui de l'omniscience. Les seules propriétés pertinentes sont celles qui résistent au regard de Dieu. On établit parfois un lien entre la cybernétique et le système de Leibniz : ceci, qui aurait peu de sens appliqué à un Wiener, par exemple, se trouve parfaitement justifié dans le cas d'Ashby. Ce dernier a décrété une fois pour toutes que le comportement d'un « système » (cerveau, organisme, etc.) avait pour seule définition rationnelle une application ou une fonction qui, à tout couple état du système-état de l'environnement, fait correspondre univoquement l'état du système à l'instant suivant. Par déduction logique, Ashby en déduit que l'auto-organisation est une notion contradictoire, qui ne participe que du domaine des apparences. Pour qu'il y ait auto-organisation, il faudrait en effet que le système puisse, de par son propre fonctionnement, déterminer le changement de la fonction qui le régit. Mais on devrait pour cela imaginer une autre fonction, d'un niveau logique supérieur à la première, qui modifierait celle-ci. Auquel cas le système aurait été défini de manière erronée par la fonction de départ. Le nouveau, la liberté, la spontanéité sont relégués au statut d'illusions ou d'erreurs. Ils sont relatifs à un observateur qui n'est surpris que parce qu'il s'est cantonné dans le point de vue fini de celui qui choisit de « ne pas soulever le couvercle du système[13] ».

Ce « théorème » d'Ashby sur l'impossibilité de l'auto-organisation fut formulé en 1962. Il devait exercer une influence considérable. Et tout d'abord sur la seconde cybernétique, qui dut s'employer à penser l'auto-organisation en contournant l'obstacle qu'il constituait. Certains des travaux les plus connus de Heinz von Foerster et de Henri Atlan s'inscrivent dans ce cadre. L'argument dont use Ashby se retrouve par ailleurs dans les domaines les plus divers, et parfois sous la plume d'auteurs qu'on imaginait mal tenir compagnie **à** la cybernétique.

L'étonnement n'est peut-être pas très grand dans le cas du sociologue allemand Niklas Luhmann, fondateur d'une théorie juridique qui repose sur le concept de système autopoiétique élaboré par Humberto Maturana et Francisco Varela. Voulant illustrer sa thèse que les systèmes normatifs sont autonomes au sens qu'ils sont clos sur eux-mêmes et qu'ils ne sauraient donc échapper à l'enclos du cercle herméneutique, Luhmann cite un texte de Douglas Hofstadter qui, de fait, ne dit rien d'autre que l'argument d'Ashby : « De quelque façon qu'un programme d'ordinateur se retourne et s'évertue à sortir de lui-même, il continuera à suivre les règles qui lui sont inhérentes [...]. Un programme peut se modifier lui-même mais il ne peut pas violer ses propres instructions – il peut tout au

mieux changer certaines de ses parties, mais cela en obéissant à ses instructions[14]. »

On éprouve une plus grande surprise à voir un philosophe peu connu pour sa culture scientifique comme Régis Debray, recourir à l'argument d'Ashby pour affirmer, au contraire si l'on peut dire, que l'autogestion est une impossiblité *logique* : « Le gouvernement d'un collectif par lui-même – *verbi gracia*, du "peuple par le peuple" – serait une opération logiquement contradictoire. » Dans sa *Critique de la raison politique*, il affirme que le « secret des malheurs collectifs [...] a la forme d'une loi logique ». Celle-ci ordonne qu'« il n'y a pas de système organisé sans clôture, et aucun système ne peut se clore à l'aide des seuls éléments intérieurs au système. La fermeture d'un champ ne peut donc procéder contradictoirement que par ouverture à un élément extérieur au champ[15] ». Même s'il attribue cette sentence à « une généralisation du théorème de Gödel » (*sic*), Debray, dans son raisonnement, ne va pas plus loin que l'argument d'Ashby : une règle ne peut se modifier que par l'opération d'une méta-règle. Chercherait-on à inclure la méta-règle dans le champ du contrôle que le système a sur lui-même que l'on devrait poser une méta-méta-règle qui, elle, échapperait à ce contrôle ; etc. On ne coupe pas à la fatalité de l'extériorité. Comme l'écrit encore Isabelle Stengers, la démarche d'Ashby « s'annonce explicitement comme devant piéger le lecteur[16] », puisqu'elle ne lui laisse d'autre choix qu'entre la soumission et l'irrationnel. Le fait est que beaucoup, malgré la mise en garde paradoxale, se sont laissés prendre au piège.

Si les archéo-cybernéticiens furent choqués par Ashby, c'est surtout, nous l'avons dit, parce qu'il introduisait l'idée radicalement nouvelle d'un réseau dont les connexions ont été tirées au hasard. Or, nouveau paradoxe, cette idée fait d'Ashby le précurseur de l'approche néoconnexionniste. Cette remarque peut faire sursauter, tant les deux démarches diffèrent dans leur esprit. Le néoconnexionniste se rapporte à son réseau comme à un objet complexe au sens de von Neumann, quasi naturel ; il en étudie expérimentalement les propriétés, le manipule, fait varier ses paramètres, etc. Ashby, lui, se sert de son automate pour fonder métaphysiquement la différence entre la vérité et l'apparence. Et cependant, il est indéniable que la propriété la plus remarquable des réseaux connexionnistes, à savoir qu'ils se stabilisent en des attracteurs qui restreignent leur dynamique à des sous-ensembles de l'ensemble des états possibles, est celle-là même sur laquelle Ashby fait fond – montrant que ce résultat ne crée la surprise que pour celui qui choisit de ne pas adopter le point de vue omniscient de l'automate sur lui-même[17] . On attribue parfois la paternité du néoconnexionnisme à la découverte *expérimentale* que fit, à la fin des années soixante, un jeune étudiant en médecine, élève de McCulloch, Stuart Kauffman : des automates booléens connectés de manière aléatoire engendrent un comportement collectif remarquable, de type cyclique et de période en général faible. Le grand biologiste C.H. Waddington se fit l'écho de cette découverte au symposium « Beyond Reductionism ». Or il en parla exactement comme Ashby l'aurait fait : tout système composé d'un très grand nombre d'éléments en interaction, affirma-t-il en substance, développe inévitablement et de façon spontanée des propriétés de stabilité[18].

Les cybernéticiens furent choqués, disions-nous, mais avaient-ils de bonnes raisons de l'être ? La rupture que représente le « cas Ashby » n'était-elle pas déjà en germe dans la démarche cybernétique elle-même ? Celle-ci s'était déjà avancée fort loin dans la voie d'une autonomisation du modèle logico-mathématique, même si elle prétendait maintenir le lien avec la réalité physique, biologique et psychologique. Elle avait déjà largement contribué à l'effort de reconstruction rationnelle et conceptuelle du donné objectai, avec

tous les risques de dérive que de telles entreprises comportent. On songe en particulier à ce qui constitue, dans l'histoire de la pensée économique, l'équivalent à peu près exact du « cas Ashby » : la décision par Gérard Debreu de reconstruire toute l'économie politique dans un cadre conceptuel de type bourbakiste. Ici aussi, pour les économistes qui sont venus après lui, il n'a plus été possible que de choisir entre se soumettre ou se voir taxer d'irrationalité.

La philosophie de l'esprit qui inspire aujourd'hui l'essentiel des sciences cognitives court les mêmes risques. Les expériences de pensée paradoxales qui lui servent de matériau à peu près exclusif, à côté des propriétés du langage ordinaire, ressemblent fort dans leur abstraction aux jeux ésotériques auxquels se sont livrés, prenant modèle sur Ashby et von Foerster, les représentants de la seconde cybernétique. Dans l'un et l'autre cas, la réflexion philosophique se nourrit davantage de science-fiction que de science. C'est pourtant au nom de la physique, la « reine des sciences », que parle la philosophie cognitive, comme l'avait fait avant elle la cybernétique. Le mot d'ordre est : « physicaliser l'intentionnalité ». Mais cette Dame Physique que les philosophes de l'esprit mettent sur un piédestal, et qui leur paraît d'autant plus pure qu'ils se gardent bien de la toucher, c'est une physique *a priori*, une physique philosophique, conceptuelle et rationnelle, et non pas la physique concrète et vivante des laboratoires. Celle-ci ressemble plus à un bricolage permanent qu'à une logique des concepts. Butant sur des problèmes, elle invente de nouvelles opérations, des procédures d'enquête, des manipulations d'objet, sans toujours se soucier de la cohérence de l'ensemble. Surtout, elle est plurielle, fragmentée. Parler de *la* physique, c'est en particulier ignorer l'« hétérogénéité entre la physique comme science des mouvements, effectivement centrée sur l'énergie, et l'ensemble des approches statistiques ou cinétiques qui définissent leur objet en termes de population et d'événements couplés[19] ». Il se trouve que cette nouvelle « science des populations », automates ou molécules, a dû abandonner une des catégories centrales sur lesquelles s'est bâtie la dynamique, thermodynamique classique comprise : la fonction d'état. Les phénomènes collectifs que cette science étudie ne sont pas intégrables. Le type de causalité auquel ils sont soumis n'est ni la causalité linéaire, ni la causalité circulaire des cybernéticiens : c'est, selon le mot d'Isabelle Stengers, une « causalité de *couplage*, au sens où la notion de couplage laisse indéterminée la nature de l'interaction (connexion, déterministe ou statistique, entre automates, réactions chimiques se produisant à l'occasion de rencontres entre molécules, diffusion de produits, etc.), pour poser la question de ce que peuvent, ensemble, une population d'entités couplées[20] ». Lorsque la philosophie cognitive, pour les raisons que nous avons analysées, s'obstine à penser l'intentionnalité sur le mode du rapport entre un *état* mental et son objet ou son contenu, elle se coupe du mouvement de la physique et de cette solidarité nouvelle qui unit le connexionnisme à la physique des systèmes désordonnés et à la thermodynamique des processus irréversibles.

A la neuvième conférence Macy, en 1952, Ross Ashby laissa les paléo-cybernéticiens abasourdis et désarmés. Leur rencontre marque, dans l'histoire de la cybernétique et, au-delà, des sciences cognitives, la fin d'une époque et le commencement d'une autre.

Les processus sans sujet

Si l'on excepte Freud, les sciences de l'homme n'ont jamais marqué beaucoup d'empressement à enraciner leurs hypothèses dans le substrat biologique et physique. Cette suffisance et cette indifférence ont parfois éveillé le désir envieux des biologistes. Jean-Pierre Changeux, aujourd'hui, cache mal sa convoitise pour ces sciences « à la mode », croit-il[21]...

Nous avons décrit la cybernétique comme une tentative physicaliste de conquérir les sciences de l'esprit, c'est-à-dire de supplanter les psychologies existantes. Il faut regretter maintenant qu'elle ait manqué totalement d'ambition en ce qui concerne les sciences de l'homme et du social en général.

Wiener déclare forfait très tôt. Ses arguments, avec le recul, ne manquent pas de décevoir. Les séries statistiques dans des conditions inchangées ne sont pas assez longues pour nous permettre d'appliquer au social les nouvelles techniques mathématiques de traitement du signal, affirme-t-il dans *Cybernetics.* Cette discipline que l'on prétend être la science des *formes* voyait vraiment la question de la société par le petit bout de la lorgnette ! Lazarsfeld ne fera pas grand chose des notions nouvelles, Jakobson ne fera que passer. Robert Merton et Talcott Parsons se montreront plus réceptifs, surtout le second, qui fera, de fait, un accueil enthousiaste à la cybernétique. Mais ce sera pour la mettre au service de son fonctionnalisme sociologique, démarche tout entière centrée sur la notion d'équilibre, et qui reste une pensée étriquée, en dépit de l'influence importante qu'elle a exercée sur la sociologie de l'après-guerre. En découvrant la cybernétique, Parsons crut y voir l'aboutissement d'un mouvement de rapprochement des sciences sociales et physiques auquel il s'était déjà préparé en suivant, dans les années trente à Harvard, en compagnie de Merton, les séminaires influents de L.J. Henderson. Celui-ci, sociologue et biochimiste, n'hésitait pas à mettre en rapport le concept d'équilibre économique de Pareto, le principe de Le Châtelier en thermodynamique et l'homéostasie de Cannon. On peut penser que l'intérêt de Parsons pour la cybernétique reposa sur un malentendu.

Certes, il y a l'anthropologie culturaliste américaine et le mouvement « personnalité et culture ». J'ai montré que le rapprochement avec la cybernétique s'était fait, là aussi, sur la base d'un énorme malentendu, qui aura présidé à l'ensemble des conférences Macy. La même chose peut être dite de l'œuvre de Gregory Bateson, originale et forte, mais en déphasage marqué, malgré qu'il en eût, avec l'esprit cybernétique. Quant aux autres, McCulloch inclus, leur culture dans les sciences de l'homme et du social était trop insuffisante pour qu'ils aperçussent le profit qu'ils auraient pu retirer à dialoguer avec elles et, *a fortiori,* à s'inspirer de leur apport.

La rencontre aurait cependant pu et dû se faire, et de façon beaucoup plus ambitieuse. J'ai insisté sur le fait que la cybernétique nous donne les moyens formels de penser la catégorie de *processus sans sujet.* Or cette catégorie, que l'on trouve aussi bien dans la philosophie sociale libérale d'un Friedrich von Hayek qu'au cœur du structuralisme marxiste de Louis Althusser, structure ce qui s'est fait de plus novateur dans la pensée du social et de l'homme de l'après-guerre. On peut en vérité distinguer deux formes bien contrastées de cette figure clé.

La première est celle de la « main invisible », ou encore de la « ruse de la Raison », qui domine entre autres la tradition libérale et économique. La question de la complexification y est centrale. On y enseigne en effet à la fois que ce sont les hommes qui font leur société, mais que celle-ci leur échappe, de la même façon qu'un véritable automate échappe à son constructeur. Or dans ce cadre, le débat qui opposa Weiss et von

Neumann à McCulloch au symposium Hixon a depuis longtemps son équivalent. L'atomisme de McCulloch s'y nomme « individualisme méthodologique », et l'on pourrait montrer que la théorie économique de l'équilibre général présente, comme le modèle du cybernéticien, cette réversibilité entre un mouvement ascendant et un mouvement descendant. Quant à la conception des totalités proposée par Weiss, ni atomiste, ni holiste, on lui trouverait aisément des équivalents, et d'abord dans l'œuvre du premier penseur de la « main invisible », Adam Smith[22]. Si les animateurs des conférences Macy avaient pris vraiment au sérieux leur projet interdisciplinaire, la mise en scène de ce débat interne aux sciences sociales eût pu les éclairer sur leur propre débat.

Certes, cette occasion manquée fut plus tard rattrapée, mais l'initiative en revint aux sciences de l'homme elles-mêmes. Le grand nom est ici celui de Friedrich von Hayek, que j'ai déjà eu l'occasion de citer plusieurs fois. Le futur Prix Nobel d'économie participa à l'une des conférences des années soixante sur l'auto-organisation réunies par von Foerster, ainsi qu'au symposium Alpbach de 1968, « Beyond Reductionism ». Ses « ordres sociaux spontanés » sont devenus, dans les colloques interdisciplinaires d'aujourd'hui, les compagnons des systèmes autopoiétiques et des structures dissipatives. Mais c'est en fait dans les ressources propres à la philosophie sociale que Hayek puisa, bien plus que dans celles de la cybernétique ou de ses avatars. L'influence fut surtout dans l'autre sens : j'ai rappelé que Rosenblatt, en concevant son Perceptron, reconnut sa dette à l'égard de Hayek.

De fait, Hayek n'avait nul besoin de chercher ailleurs que dans la tradition libérale des « Lumières écossaises » de quoi penser ce qui est l'équivalent, dans l'ordre du social, de la conjecture de von Neumann sur la complexité. Cette tradition tient en effet pour non contradictoires les deux propositions suivantes (configuration analogue à celle que nous avons présentée plus haut au sujet de la vie et des processus physico-chimiques qui la produisent) : 1) ce sont les hommes qui « agissent » leur société ; 2) la société leur échappe, parce qu'elle est (infiniment) plus complexe qu'eux. On peut sans incohérence être non réductionniste sans avoir à accepter le holisme. S'il est vrai qu'il s'agit encore d'un individualisme méthodologique, il s'oppose à ce que l'on entend habituellement par là, précisément par le refus du réductionnisme. On parlera d'individualisme méthodologique *complexe.*

Pour Hayek, l'« ordre social spontané » constitue un troisième type d'ordre, à côté de l'ordre naturel et de l'ordre artificiel. C'est une émergence, un effet de composition, un effet de système. Le « système » n'est évidemment pas un sujet doué de conscience et de volonté. Le savoir qu'il mobilise est irréductiblement distribué sur l'ensemble de ses éléments constitutifs, il ne saurait être synthétisé en un lieu qui serait celui du « savoir absolu » du système sur lui-même. Ce savoir collectif est porté par l'ordre social en tant qu'il est le « produit de l'action des hommes mais non de leurs desseins », et n'est récapitulable par aucune conscience individuelle. C'est un *savoir sans sujet.* Il s'incarne dans des normes, des règles, des conventions, des institutions, lesquelles s'incorporent dans les esprits individuels sous la forme de schèmes abstraits. Comme l'écrit Hayek : « L'esprit ne fabrique point tant des règles qu'il ne se compose de règles pour l'action [...]. Nous avons à notre service une si riche expérience, non parce que nous possédons cette expérience mais parce que, sans que nous le sachions, elle s'est incorporée dans les schémas de pensée qui nous conduisent[23]. »

Au symposium « Beyond Reductionnism », Hayek fut amené à penser ce rapport

paradoxal entre un savoir collectif sans sujet et les sujets individuels en termes d'« autotranscendance » (*self-transcendence*). On ne peut manquer de faire le rapprochement avec la figure phénoménologique de la « transcendance dans l'immanence » – également ratée par la cybernétique, comme on l'a vu. Dans le cas de l'esprit, comme dans celui du social, la figure paradoxale, mais que l'on peut désormais penser en termes de *mécanismes*, est celle d'une mise en extériorité de soi par rapport à soi.

J'ai déjà évoqué, dans le chapitre 4, l'autre forme que prend la figure des processus sans sujet. Elle relève de ce « nietzschéo-heideggérianisme » qui va informer les sciences de l'homme de l'après-guerre dans leur réaction contre les philosophies de la conscience ; et ce surtout en France, avec les vagues successives du structuralisme, du post-structuralisme et de la « déconstruction ». La cybernétique, j'ai tenté de le montrer, faisait plus qu'annoncer ce moment de « désanthropologisation » et de déshumanisation de la figure de l'homme. On eût pu souhaiter qu'elle en prît… conscience, en cherchant le dialogue avec ses compagnons de route, même distants. Cela ne fut pas. Dans ce cas aussi, le chemin se fit dans l'autre sens. J'ai évoqué le cas de Lévi-Strauss, et celui de Lacan. On peut certes penser que les résultats que le structuralisme crut pouvoir établir au moyen des outils cybernétiques ne résistent pas à l'analyse critique : la démonstration de la stérilité de l'« imaginaire » – c'est-à-dire de ce que nous avons appelé, à propos des expériences de Bavelas, la « spécularité » –, la preuve de l'existence d'un niveau « symbolique » en surplomb, structuré comme un langage et fonctionnant comme une machine[24]. Tout cela a fortement vieilli. On peut aussi penser que la transcendance postulée du symbolique par rapport à l'imaginaire manque précisément la figure de l'auto-transcendance propre à l'autre tradition. On ne peut en tout état de cause nier qu'il y ait eu des affinités profondes entre le projet cybernétique et la déconstruction de la conception métaphysique du sujet.

Le rendez-vous manqué avec les sciences de l'homme

En conclusion, je voudrais brièvement évoquer la manière dont je vois les rapports entre les sciences sociales et les sciences cognitives aujourd'hui.

L'individualisme méthodologique en sciences sociales a une règle d'or : *ne jamais traiter les collectifs comme des sujets.* Ce serait une faute sur les catégories. Or cette faute, nous sommes constamment tentés de la commettre. Considérons l'élection ou le vote, qui sont aujourd'hui la procédure démocratique par excellence. On peut songer à l'élection d'un président au moyen du suffrage universel, ou bien à un référendum où l'on doit répondre par oui ou par non à une certaine question, etc. Il arrive fréquemment que la répartition des suffrages soit très proche de l'égalité. Le vainqueur l'emporte de très peu. Tant et si bien qu'une distribution des votes massivement en faveur de l'une des options fait inévitablement suspecter que la consultation ne s'est pas faite de manière régulière – de la même façon qu'une série de tirages à pile ou face qui donne à pile sensiblement plus, ou moins, que la moitié des coups provoque immanquablement le soupçon que la pièce est truquée. Depuis Montesquieu, la théorie de la démocratie a multiplié les explications pour rendre compte de cette observation. Pour un théoricien de l'information,

cependant, le résultat du vote est tout simplement celui qui maximise l'entropie, c'est-à-dire le désordre. La procédure est formellement équivalente à un tirage au sort.

Cette interprétation a de quoi choquer, mais elle prend encore plus de poids si l'on est sensible à ce que la théorie de la décision rationnelle nomme le « paradoxe du vote ». Sauf dans le cas extrêmement improbable où les suffrages se répartissent également entre les deux options, il est inévitable de dire, dans une perspective individualiste, que le bulletin déposé dans l'urne par chacun des électeurs aura eu un effet strictement nul. A la question : « Le résultat final eût-il été changé si j'avais voté autrement que je l'ai fait », chacun doit répondre : non ! Or qu'observe-t-on ? Le résultat du vote est souvent interprété comme la manifestation du choix soigneusement délibéré d'un sujet collectif : le peuple, l'électorat, etc. La consultation a donné l'avantage au oui, mais d'extrême justesse. On interprète : « Dans sa grande sagesse, le peuple a répondu oui à la question posée, mais il a aussi voulu donner un avertissement à tous ceux qui voulaient précipiter les événements… » Tout se passe donc comme si l'on faisait du hasard un sujet, et un sujet collectif en situation d'extériorité, puisque, évidemment, le peuple est censé transcender chacun des citoyens pris individuellement. Dans la vie des innombrables comités ou commissions auxquels les sociétés modernes confient le soin de l'administration des choses publiques, le recours au vote anonyme n'est bien souvent qu'un moyen déguisé de déléguer au hasard la décision que la discussion à coups d'arguments supposés rationnels s'est montrée incapable d'atteindre. Mais ces formes d'engendrement du hasard sont considérées comme légitimes et porteuses de sens, dans l'exacte mesure où elles sont productrices d'extériorité ou de transcendance, et où elles peuvent être tenues pour les décisions d'un sujet collectif.

Cette attitude interprétative, nous explique le philosophe de l'esprit Daniel Dennett, est en vérité inévitable. Nous attribuons sans cesse à autrui des « états mentaux » (intentions, désirs, croyances, etc.), que cet autrui soit un autre être humain, un animal, une machine… ou un collectif humain. C'est ce que Dennett appelle la « posture intentionnelle » (*intentional stance*[25]).

Cette posture conduit pratiquement à affaiblir-je préfère dire : complexifier – l'individualisme méthodologique, puisque le sujet individuel n'a plus le monopole de certains attributs de la subjectivité. Il faut admettre qu'à côté de ces sujets individuels il existe des *quasi-sujets,* qui sont des entités collectives capables d'exhiber certains au moins des attributs que l'on croyait réservés aux « véritables » sujets – les individus – et, en particulier, l'existence d'états mentaux. On n'hésitera pas ainsi à dire d'une organisation, et plus généralement d'une entité collective, qu'elle est capable d'apprendre, mais aussi de savoir, de se souvenir, d'analyser une situation, de faire des expériences, de former des concepts, de prendre des décisions et d'agir.

Or, de leur ancêtre cybernétique aux développements actuels, les sciences cognitives nous présentent le sujet individuel lui-même comme un *quasi-sujet,* c'est-à-dire comme un collectif manifestant les propriétés de la subjectivité. Lorsque je pense, je me souviens, je désire, je crois, je décide, etc., ce n'est pas un fantôme dans la machine cérébrale, un homunculus caché, qui est le sujet de ces prédicats – c'est la machine elle-même, sous la forme, par exemple, d'un réseau de neurones. Il n'y a pas de fantôme dans la machine, il n'y a pas de centre de la subjectivité, comme y insiste Dennett. Les attributs de la subjectivité sont des effets émergents produits par le fonctionnement spontané, « auto-organisé », d'une organisation complexe en forme de réseau. Les spécialistes des sciences

cognitives qui défendent ce genre de thèses utilisent parfois des expressions étranges pour les dire. Varela parle de *selfless selves*[26] et Dennett d'« egos non égoïques[27] ».

En d'autres termes : l'affaiblissement, voire la « déconstruction » de la conception métaphysique, c'est-à-dire cartésienne et leibnizienne, de la subjectivité s'opère dans le champ commun aux sciences sociales et aux sciences cognitives à la fois par en haut et par en bas. Par en haut : les attributs de la subjectivité ne sont pas le monopole des sujets individuels. Des entités collectives peuvent les manifester aussi bien. Par en bas : les attributs de la subjectivité ne sont pas les attributs d'un sujet, ce sont des effets émergents produits par le fonctionnement de processus sans sujet.

Dans les deux cas, l'outil de cette déconstruction est le même, c'est la compréhension qu'un réseau complexe d'interactions entre des entités simples – neurones formels pour le quasi-sujet individuel, individus schématiques pour le quasi-sujet collectif – peut exhiber des propriétés remarquables. Pour un Daniel Dennett, il n'est *ni plus, ni moins* fondé d'attribuer un état mental, par exemple une intention, à un être humain qu'à une collectivité.

On observe donc qu'il n'est nul besoin de la « déconstruction » philosophique patentée pour déconstruire la métaphysique de la subjectivité. Le mouvement des sciences, n'en déplaise aux heideggériens, va plus loin dans ce sens qu'il n'ont jamais osé le rêver.

Qui a écrit ceci ?

« De la même façon que les araignées n'ont pas à penser à la façon dont elles tissent leur toile, de même nous, les êtres humains, n'avons pas à déterminer consciemment et volontairement quels récits nous avons à raconter, et comment les raconter, afin de constituer notre moi : nos récits sont certes tissés (ce sont des textures, des *textes*), mais pour l'essentiel, ce n'est pas nous qui les tissons. La conscience humaine, c'est-à-dire le soi en tant qu'il est récit, est le produit de ces textes, elle n'en est pas la source. »

On pourrait croire du Derrida. C'est le texte qui nous fait, ce n'est pas nous qui faisons le texte. En vérité, il s'agit d'un passage, librement traduit, du livre de Daniel Dennett consacré à la (déconstruction de la) conscience[28].

Qui lit le philosophe cognitif sur la conscience et le soi ne peut qu'être frappé par son recours constant à des métaphores sociales et politiques, et mêmes électorales. L'émergence d'un soi à un moment donné, au cours d'un de ces processus qui constituent la vie de l'esprit, c'est comme le passage de l'état de foule – foule d'événements mentaux, de configurations neuronales – à celui de communauté politique organisée, se dotant par élection, non d'un centre de contrôle, mais d'un représentant : le *Head of Mind* joue le même rôle que le *Head of State.* Dans ce scénario, il y a parfois des crises : plusieurs représentants potentiels sont en rivalité, ils se disputent le pouvoir, etc. Les mécanismes mentaux et les mécanismes sociaux sont mis sur le même plan, en tant précisément qu'ils *sont* des mécanismes.

La cybernétique annonçait tout cela, était porteuse de tout cela, mais elle n'a pas su le voir parce qu'il eût fallu qu'elle consentît à s'intéresser à ce que des disciplines établies, et parfois depuis longtemps, avaient à dire sur son sujet : la vie de l'esprit. Sans doute était-elle trop impatiente, trop pénétrée de ce qu'elle pensait être sa mission révolutionnaire, trop sûre de représenter l'avant-garde de la science en marche vers des terres nouvelles. Elle se trouve aujourd'hui bien oubliée ou bien mal aimée, injustement. Mais les sciences cognitives sont là qui, encore plus sûres d'elles-mêmes, projettent à leur tour de reconstruire la science de l'homme en faisant table rase de tout ce qui s'est pensé

jusqu'à elles. Si l'histoire, héroïque et malheureuse, de la cybernétique doit nous apprendre quelque chose, c'est probablement que, à côté de l'esprit de pionnier, la modestie, le doute raisonné et l'attention, nourrie d'esprit critique, à la tradition sont des vertus indispensables à l'aventure de la connaissance.

Index *

Notes

Avant-propos – Note 1

J.-P. DUPUY, *Ordres et Désordres*, Seuil, Paris, 1982, « Introduction ».

— Retour au texte —

Avant-propos – Note 2

P. Dumouchel et J.-P. Dupuy, *L'Auto-organisation : de la physique au politique*, Seuil, Paris, 1983.

— Retour au texte —

Avant-propos – Note 3

A l’époque, le sigle CRÉA signifiait : « Centre de recherche épistémologie et autonomie ».

— Retour au texte —

Avant-propos – Note 4

MIT Press, Cambridge (Mass.), 1980.

— Retour au texte —

Avant-propos – Note 5

In *Cahiers du CRÉA*, n° 7, novembre 1985.

— Retour au texte —

Avant-propos – Note 6

The MIT Press.

— Retour au texte —

1 – Note 1

Herbert SIMON, *Models of my Life*, Basic Books, 1991, p. 305.

— Retour au texte —

1 – Note 2

Hannah ARENDT, *Condition de l’homme moderne*, Calmann-Lévy, Paris, 1961, p. 336.

— Retour au texte —

1 – Note 3

Ihid., p. 333.

— Retour au texte —

1 – Note 4

Ibid., p. 332.

— Retour au texte —

1 – Note 5

Jean ULLMO, *La Pensée scientifique moderne*, Flammarion. Paris, 1969, p. 99. C'est ce qu'il appelle la « structure » que Jean Ullmo caractérise ainsi. Mais plus loin (p. 104), il identifie modèle et structure. On peut attribuer à l'influence que le structuralisme exerçait à l'époque le choix d'un mot plutôt que l'autre.

— Retour au texte —

1 – Note 6

Jean ULLMO. *La Pensée scientifique moderne, op. cit.*, p. 104.

— Retour au texte —

1 – Note 7

Jean ULLMO, *La Pensée scientifique moderne, op. cit.,* p. 116-117.

— Retour au texte —

1 – Note 8

Prenant ses distances avec la lecture par Hannah Arendt du *verum factum,* Philippe RAYNAUD insiste sur ce point dans son *Max Weber et les dilemmes de la raison moderne*, PUF, Paris, 1987, 2e section, chap. 1 : « L'idée de science de l'Esprit ».

— Retour au texte —

1 – Note 9

Chap. x, 5, A. Blanchard, 1974, p. 146.

— Retour au texte —

1 – Note 10

Proc. London Math. Soc., p. 42-43 ; repris *in* M. DAVIS (ed.), *The Undecidahle,* Hewlett, New York, Raven Press, 1965.

— Retour au texte —

1 – Note 11

Monatshefte für Mathematik und Physik, 38 ; trad. anglaise *in* DAVIS, *op. cit.*

— Retour au texte —

1 – Note 12

A. CHURCH, « An Unsolvable Problem of Elementary Number Theory », *American Journal of Mathematics,* vol. 58, 1936 ; repris *in* DAVIS, *op. cit.*

— Retour au texte —

1 – Note 13

J. MOSCONI, *La Constitution de la théorie des automates,* thèse de doctorat d'État, Université Paris-I, 1989 ; t. 1, p. 19.

— Retour au texte —

1 – Note 14

A. TURING, « Computability and Lambda-definability », *Journal of Symbolic Logic*, 2, 1937.

— Retour au texte —

1 – Note 15

J. MOSCONI, « Sur quelques capacités et incapacités des machines », *Bulletin de la Société française de philosophie*, 3, juillet-septembre 1991, p. 86.

— Retour au texte —

1 – Note 16

La question : « Le monde est-il récursif ? » agite certains philosophes et logiciens. Cf, en France, les travaux de Jean-Paul DELAHAYE et de Jean-Pierre DUBUCS.

— Retour au texte —

1 – Note 17

Ainsi, Hilary PUTNAM : cf. l'évolution de ses idées dans *Philosophical Papers, II*, Cambridge University Press, 1975.

— Retour au texte —

1 – Note 18

Mind, LIX, 1950 ; repris *in* A. R. ANDERSON(ed.), *Minds and Machines* ; trad. française de ce dernier recueil : *Pensée et Machines*, Champ Vallon, Seyssel, 1983.

— Retour au texte —

1 – Note 19

Dans la thèse qu'il fait actuellement au CRÉA sous la direction de Daniel Andler, Jean Lassègue déroule de manière très subtile toutes les implications sexuelles de ce dispositif, en les mettant en rapport avec la biographie d'Alan Turing. cf. J. LASSÈGUE, « Le test de Turing et l'énigme de la différence des sexes » in D. ANZIEU et al. (eds.). *Les contenants de pensée*, Dunod, Paris, 1993, p. 145-195.

— Retour au texte —

1 – Note 20

Allusion au dispositif théorique mis en place par John RAWLS pour bâtir sa *Théorie de la justice* (Seuil, Paris, 1987).

— Retour au texte —

1 – Note 21

J. SEARLE, « Minds, Brains and Programs », *Behavioral and Brain Sciences*, vol. 3, 3, 1980, p. 417-458 ; voir aussi J. SEARLE, « L'esprit est-il un programme d'ordinateur ? », *Pour la Science*, 149, mars 1990, p. 38-45.

— Retour au texte —

1 – Note 22

Cité par J. MOSCONI, « Sur quelques capacités et incapacités des machines », *loc. cit.,* p. 97.

— Retour au texte —

1 – Note 23

Certains travaux de psychologie du développement semblent montrer que la capacité de *faire semblant* ou de *mimer* est liée à la faculté de projeter ses propres états mentaux sur autrui et de *simuler* les états d'esprit d'autrui à partir des siens propres, et constitue une étape précoce et fondamentale du développement cognitif de l'enfant ; cf. A. GOLDMAN, « Interpretation psychologized », *Mind and Language,* 4, 3, 1989, p. 161-185. Pour une discussion critique de cette thèse, voir P. ENGEL, *États d'Esprit*, Alinéa, Aix-en-Provence, 1992, chap. 3 : « Théories de l'interprétation et théorie de l'esprit » ; réédition *Introduction à la philosophie de l'esprit*, La Découverte, Paris, 1994.

— Retour au texte —

1 – Note 24

Allusion à la polémique qui a opposé les deux philosophes dans la revue *Glyph*, en 1977, à propos de la théorie d'Austin sur les actes de langage.

— Retour au texte —

1 – Note 25

V. DESCOMBES, *Le Même et l'Autre. Quarante-cinq ans de philosophie française (1933-1978),* Minuit, Paris, 1979. Descombes se réfère à l'« Essai sur la pensée d'Emannuel Levinas » qui constitue le chapitre 4 de *L'Écriture et la Différence*, Seuil, Paris. 1967, p. 133. A ceci près que Descombes triche puisqu'il y est question de feindre de parler… *grec* ! Ce passage au chinois est un coup de maître si l'on songe que le texte de Descombes date de 1979 et que ce n'est que l'année suivante que Searle mettra au point l'argument de la pièce chinoise…

— Retour au texte —

2 – Note 1

Fin de non-recevoir adressée par Marvin Minsky à Françoise Fogelman-Soulié, venue l'interviewer pour le compte de la recherche sur l'histoire des théories de l'auto-organisation (voir Avant-propos). MIT, juillet 1983.

— Retour au texte —

2 – Note 2

Cf. S. Heims, *John von Neumann and Norbert Wiener, op. cit*., p. 183-184.

— Retour au texte —

2 – Note 3

Cf. l'introduction de A. BURKS à son édition de *Theory of Self Reproducing Automata* de John VON NEUMANN, University of Illinois Press, 1966.

— Retour au texte —

2 – Note 4

Cf. S. HEIMS, « Gregory Bateson and the Mathematicians : From interdisciplinary interaction to societal functions », *Journal of the History of the Behavioral Sciences*, 13, 1977, p. 141-159 (p. 143) ; et S. HEIMS, *The Cybernetics Group, op. cit.,* p. 289, note 6.

— Retour au texte —

2 – Note 5

H. VON FOERSTER, « Molecular Ethology », 1970, repris *in Observing Systems*, Intersystems Publications, 1981.

— Retour au texte —

2 – Note 6

Par exemple, Francisco VARELA, *Principles of Biological Autonomy*, Elsevier North Holland, 1979 (trad. française par P. Bourgine et P. Dumouchel : *Autonomie et Connaissance*, Seuil, Paris, 1989).

— Retour au texte —

2 – Note 7

Ph. Breton, « La cybernétique et les ingénieurs dans les années cinquante », *Culture technique*, n° 12, mars 1984, p. 158.

— Retour au texte —

2 – Note 8

Macy 8, p. 168-171

— Retour au texte —

2 – Note 9

« Why the Mind is in the Head », communication au symposium Hixon, 1948. Repris dans *Embodiments of Mind*, The MIT Press, 1965.

— Retour au texte —

2 – Note 10

Au symposium Hixon (*Hixon,* p. 68-69).

— Retour au texte —

2 – Note 11

Steve HEIMS a significativement intitulé l'un de ses articles « Encounter of Behavioral Sciences with new machine-organism analogies in the 1940's », *Journul of the History of the Behavioral Sciences*, 11,1975, p. 368-373.

— Retour au texte —

2 – Note 12

« *Mysterium Iniquitatis* of Sinful Man Aspiring into the Place of God », repris *in Embodiments of Mind, op. cit.,* p. 163.

— Retour au texte —

2 – Note 13

Hixon, p. 32

— Retour au texte —

2 – Note 14

Titre d'une conférence donnée en 1961 (« Qu'est-ce qu'un nombre, pour qu'un homme puisse le connaître, et un homme, pour qu'il puisse connaître un nombre ? ») et qui résume bien la quête philosophique qui, du propre aveu de McCulloch, a donné sens à toute son œuvre. Repris *in Embocliments of Mind, op. cit*., p. 8.

— Retour au texte —

2 – Note 15

Lors d’une confession autobiographique à l’occasion d’un colloque organisé à Montpellier, en janvier 1984, par Jean-Louis Le Moigne, sur son œuvre.

— Retour au texte —

2 – Note 16

Jean-Pierre CHANGEUX, *L'Homme neuronal,* Fayard, Paris, 1983, p. 125.

— Retour au texte —

2 – Note 17

Ibid., p. 52. Je souligne.

— Retour au texte —

2 – Note 18

Sur l'histoire de ces dispositifs, que l'on nomme généralement « machines logiques » – mais dans un sens différent de celui dans lequel j'emploie moi-même l'expression dans le corps du texte, puisqu'il s'agit de machines matérielles –, on peut consulter l'ouvrage classique de Martin GARDNER, *Logicai Machines and Diagrams*, Mc Graw Hill, New York, 1959.

— Retour au texte —

2 – Note 19

Jean MOSCONI met clairement ce point en évidence dans *La Constitution de la théorie des automates, op. cit.,* t. I., p. 59-62.

— Retour au texte —

2 – Note 20

Macy 9, p. 147-148.

— Retour au texte —

2 – Note 21

Cf. le témoignage de Norman Geschwind in *McC. Intro*. Sur la guerre qui, de 1870 à 1950 et au-delà, opposa les « réticularistes » (continuistes) aux « neuro-nistes », voir l'analyse très détaillée de Jean-Pierre CHANGEUX *in L'Homme neuronal, op. cit.,* chap. 1. Au symposium Hixon (1948), Lashley, dans la discussion qui suit l'exposé de McCulloch, « Why the Mind is in the Head », recourt à des arguments qui rappellent fort ceux que Pierre Flourens opposait à Gall, en 1842, plus de cent ans auparavant. (Cf. *L'Homme neuronal*, p. 28, et *Hixon*, p. 70-71).

— Retour au texte —

2 – Note 22

Dans le modèle initial de McCulloch et Pitts, les coefficients de pondération – que l'on nomme aujourd'hui « poids synaptiques » – sont tous égaux à un ; de telle sorte qu'un neurone est mis à feu si et seulement si le nombre des impulsions qu'il reçoit est supérieur au seuil.

— Retour au texte —

2 – Note 23

Dans sa description des mécanismes qui expliquent ce qu'il appelle le « passage à l'acte », Jean-Pierre Changeux reprend ce même schéma. La seule différence, c'est que l'organe élémentaire de calcul n'est plus aujourd'hui identifié au neurone, mais à sa membrane. Comme Changeux lui-même le fait remarquer, à chaque étape de l'histoire des schémas explicatifs, « il y a "réduction" d'un niveau d'organisation à un niveau plus élémentaire ». Cf. *L'Homme neuronal*, *op. cit.,* p. 132 et 123.

— Retour au texte —

2 – Note 24

Brain, 53-Part 2, 166, 1930. Le titre signifie : « Application théorique à certains problèmes neurologiques des propriétés d'ondes excitatrices se déplaçant à l'intérieur de circuits fermés ».

— Retour au texte —

2 – Note 25

Lawrence KUBIE, « Repetitive Core of Neuroses », *Psychoanalytic Quarterly*, 10,23 (1941). J. -P. CHANGEUX, dans *L'Homme neuronal* (cf. p. 178-179 et 211), semble ignorer le rôle déterminant de Kubie dans la formation de cette hypothèse.

— Retour au texte —

2 – Note 26

Cf. l’échange entre McCulloch et Kubie dans *Macy* 6, p. 61-62.

— Retour au texte —

2 – Note 27

J. MOSCONI, *La Constitution de la théorie des automates*, *op. cit.,* t. I, p. 154.

— Retour au texte —

2 – Note 28

J. VON NEUMANN, « Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components », conférences au CalTech (Institut californien de technologie), janvier 1952. Pour le point de vue de McCulloch sur l'enchaînement de ces idées, voir *Embodiments of Mind, op. cit.,* p. 11.

— Retour au texte —

2 – Note 29

Cf. sa conférence « Finality and Form in Nervous Activity », prononcée en 1946, publiée en 1952, reprise dans *Embodiments of Mind, op. cit.,* p. 272-273.

— Retour au texte —

2 – Note 30

Il est fort bien analysé par J. MOSCONI dans son ouvrage déjà cité, p. 137-138.

— Retour au texte —

2 – Note 31

Beaucoup, à l'époque, comparaient les capacités intellectuelles de Pitts à celles de von Neumann. On a dit que c'était le seul qui pouvait comprendre l'ensemble de ce qui se discutait aux conférences Macy.

— Retour au texte —

2 – Note 32

Chicago, University of Chicago Press, vol. 5, 1943. L'article est plus facilement accessible dans le recueil *Embocliments of Mind, op. cit.,* p. 19-39.

— Retour au texte —

2 – Note 33

S.C. Kleene, « Representations of events in nerve nets and finite automata », *in* C.E. SHANNON et J. MCCRTHY (eds) *Automata Studies*, Princeton University Press, 1956, p. 3-42.

— Retour au texte —

2 – Note 34

Voir la conférence de 1964 : « What’s in the Brain That Ink May Character ? », reprise dans *Embodiments of Mind, op. cit.,* p. 387-397 (spécialement p. 393).

— Retour au texte —

2 – Note 35

Voir la conférence de 1948, « Through the Den of the Metaphysician » ; reprise dans *Embodiments of Mind, op. cit.,* p. 142-156 (citation p. 143-144). Il existe une version française de cette conférence, parue dans *Thadès*, vol. 7, PUF., 1951, p. 37-49, sous le titre : « Dans l'antre du métaphysicien ». La traduction dans le corps du texte est de moi.

— Retour au texte —

2 – Note 36

« What’s in the Brain That Ink May Character ? », *loc. cit.,* p. 389 et p. 393.

— Retour au texte —

2 – Note 37

Cf. par exemple la discussion avec von Neumann au symposium Hixon (*Hixon,* p. 58-61) ; et J. Y. LETTVIN, « Warren and Walter », *in McC. Intro*, p. 5 et 13.

— Retour au texte —

2 – Note 38

Repris *in Theory of Self-Reproducing Automata*, University of Illinois Press, 1966. Je souligne.

— Retour au texte —

2 – Note 39

L'Homme neuronal, op. cit., p. 122-123.

— Retour au texte —

2 – Note 40

Ibid., p. 211.

— Retour au texte —

2 – Note 41

Macy 6, p. 12-26 et 60-61.

— Retour au texte —

2 – Note 42

Alors même qu'il comporte une analyse historique très fouillée, le livre de Jean-Pierre CHANGEUX, *L'Homme neuronal, op. cit.,* ne mentionne pas une seule fois le nom de McCulloch. Quant à Marc JEANNEROD, *Le Cerveau-machine*, Fayard, Paris, 1983, il nous plonge dans la confusion la plus totale puisque, si l'auteur fait bien une référence furtive et approximative au réseau de neurones à seuil, c'est pour en attribuer la paternité aux « travaux de Eccles, dans les années 1940-1950 » (*sic*).

— Retour au texte —

2 – Note 43

Cf. Jerome Y. LETTVIN, « Introduction to vol. I » (p. 1 et 2) et « Warren and Walter » (p. 12), *in McC. Intro*.

— Retour au texte —

2 – Note 44

Le livre de H. GARDNER, *The Mind's New Science*, Basic Books, New York, 1985, est surtout anecdotique.

— Retour au texte —

2 – Note 45

Frank ROSENBLATT, *Principles of Neurodynamics. Perceptrons and the Theory of Drain Mechanisms*, Spartan Books, 1962.

— Retour au texte —

2 – Note 46

Ibid., p. 16-20.

— Retour au texte —

2 – Note 47

Cambridge, The MIT Press, 1969.

— Retour au texte —

2 – Note 48

The MIT Press, 1969.

— Retour au texte —

2 – Note 49

Hubert Dreyfus, *What Computers Can't Do*, 2e éd., Harper and Row, 1979, p. 130.

— Retour au texte —

2 – Note 50

Daniel MEMMI, « Connexionnisme, Intelligence Artificielle et modélisation cognitive », *Intellectica*, n°spécial consacré aux « Modèles connexionnistes », n° 9-10, 1990. Sur cette même question, on consultera aussi avec profit Daniel ANDLER : « Introduction – Calcul et représentation : les sources », *in* D. ANDLER (éd.). *Introduction aux sciences cognitives,* Folio essais, Gallimard, Paris, 1992 ; et, du même auteur, « Connexionnisme et cognition : A la recherche des bonnes questions », *Revue de synthèse* : IVe S. n° 1-2, janvier-juin 1990, p. 95-127.

— Retour au texte —

2 – Note 51

Plus précisément, Rosenblatt a utilisé l'adjectif *connectionist.* Mais Donald HEBB l'avait déjà fait dans son livre *The Organization of Behavior*, publié en 1949. C'est dans cet ouvrage qu'il proposait une règle d'apprentissage restée célèbre, et qui est toujours jugée pertinente, dans les réseaux de neurones : une connexion est renforcée si elle relie deux neurones qui sont mis à feu en même temps. Pour une anthologie des textes fondateurs du connexionnisme, voir J. A. ANDERSON et E. ROSENFELD (eds). *Neurocomputing. Foundations of Research*. The MIT Press. 1988.

— Retour au texte —

2 – Note 52

Cf. Daniel ANDLER, « From paleo to neo-connectionism », *in* G. VAN DER VIJVER (ed.). *New Perspectives on Cybernetics*, Dordrecht. Kluwer, 1992. p. 125-146.

— Retour au texte —

2 – Note 53

Sur cet historique, voir l'« Introduction » d'Arthur Burks à son édition des travaux de von Neumann sur la théorie des automates : *Theory of Self Reproducing Automata, op. cit.,* p. 6-12 ; et le classique H. H. GOLDSTINE, *The Computer from Pascal to von Neumann*, Princeton University Press, 1972. Von Neumann a consigné ses réflexions dans son rapport jamais publié : « First Draft of a Report of the EDVAC », en juin 1945 (cependant, on en trouvera des extraits dans B. Randell (éd.), *The Origins of Digital Computers*, Springer-Verlag, Berlin, New York, 3°éd., 1982).

— Retour au texte —

2 – Note 54

S. HEIMS, « Encounter of Behavioral Sciences with new machine-organism analogies in the 1940's », *loc. cit.,* p. 371.

— Retour au texte —

2 – Note 55

Hixon, p. 22-24.

— Retour au texte —

2 – Note 56

Voir l'« Introduction » de BURKS a *Theory of Self-Reproducing Automata, op. cit.*, p. 19 et 25-28.

— Retour au texte —

2 – Note 57

Voir sur ce point Jerome LETTVIN, « Introduction to Vol. I », p. 16-17, *in McC. Intro.*

— Retour au texte —

2 – Note 58

Daniel DENNETT, « The Logical Geography of Computational Approaches – A View from the East Pole », MIT Sloan Conference, mai 1984.

— Retour au texte —

3 – Note 1

G. BATESON, *Steps to an Ecology of Mind*, Ballantine Books, 1972. Citation extraite d’une conférence prononcée en 1966.

— Retour au texte —

3 – Note 2

Source : « A Note by the Editors », *in Macy* 8, p. XIX. BATESON se trompe lorsqu'il affirme, dans le « Foreword » de *Steps to an Ecology of Mind, op. cit.,* que Bigelow était également présent (p. IX).

— Retour au texte —

3 – Note 3

Source : Wiener, *Cybernetics, op. cit.,* p. 12.

— Retour au texte —

3 – Note 4

Source : Communication de Fremont-Smith à Steve Heims, *in John von Neumann and Norbert Wiener, op. cit.,* note 13, p. 468.

— Retour au texte —

3 – Note 5

Thèmes abordés par les cycles de conférences de la fondation en 1949. Source : *Macy* 6, p. 9.

— Retour au texte —

3 – Note 6

S. HEIMS, *op. cit.,* p. 185-186. WIENER se trompe, p. 15 de son « Introduction » à *Cybernetics*, en situant cette rencontre « in the late winter of 1943-1944 ».

— Retour au texte —

3 – Note 7

Rosenblueth, retenu à Mexico, était absent. Voir S. HEIMS, « Gregory Bateson and the Mathematicians », *loc. cit.,* p. 143, n° 11 ; et N. WIENER, *Cybernetics,* p. 15.

— Retour au texte —

3 – Note 8

Lettre de Wiener à Rosenblueth, 1945, citée par S. HEIMS, *John von Neumann and Norbert Wiener, op. cit.,* p. 186.

— Retour au texte —

3 – Note 9

S. Heims, *op. cit.,* p. 189.

— Retour au texte —

3 – Note 10

Source : S. HEIMS, *John von Neumann and Norbert Wiener, op. cit.,* p. 202 ; « Gregory Bateson and the Mathematicians », *loc. cit.,* p. 145 ; *The Cybernetics Group, op. cit.,* p. 24-26.

— Retour au texte —

3 – Note 11

Voir S. HEIMS, « An Encounter between neo-mechanists and the Human Sciences », *Cahiers du CRÉA,* n° 7, octobre 1985.

— Retour au texte —

3 – Note 12

S. HEIMS, « Gregory Bateson and the Mathematicians », *foc. cit.,* p. 151.

— Retour au texte —

3 – Note 13

S. Heims, « Mechanists and social "scientists" (1946-1953) », *loc. cit. ;* et *The Cyberneties Group, op. cit.,* p. 183.

— Retour au texte —

3 – Note 14

Source : « A Note by the Editors », *in Macy* 8, p. XIX.

— Retour au texte —

3 – Note 15

Tel. Mech., p. 192.

— Retour au texte —

3 – Note 16

La conférence de Wiener s’intitulait « Time, Communication, and the Nervous System ».

— Retour au texte —

3 – Note 17

« A note by the Editors », *in Macy* 8, p. XIX-XX.

— Retour au texte —

3 – Note 18

Cf. l'« Introduction » de WIENER à *Cybernetics, op. cit.,* p. 11-12. Wiener ne mentionne pas l'usage antérieur (et bien différent) de ce mot par Ampère, mais fait référence au « governor » de Maxwell (1868).

— Retour au texte —

3 – Note 19

« A note by the Editors », *in Macy* 8, p. XIX-XX.

— Retour au texte —

3 – Note 20

Von Foerster raconte cette histoire dans sa contribution aux introductions aux œuvres complètes de McCulloch : « Circular Causality – The Beginnings of an Epistemology of Responsibility » (*McC. Intro.*).

— Retour au texte —

3 – Note 21

Voir le témoignage de Jerome LETTVIN(qui accompagna McCulloch, en 1952, dans son transfert de Chicago à Cambridge) dans « Warren and Walter », p. 8, *in McC. Intro.* Voir aussi Steve HEIMS, *John von Neumann and Norbert Wiener, op. cit.,* p. 190, n° 34. Il faut aussi noter la rivalité croissante entre Wiener et von Neumann (cf. HEIMS, *ibid.,* p. 208). Voir aussi *The Cybernetics Group*, *op. cit.,* p. 137-140.

— Retour au texte —

3 – Note 22

Cette lettre, signée de McCulloch, s'intitule « An Account of the first three Conferences on Teleological Mechanisms ». Archives personnelles de Heinz von Foerster.

— Retour au texte —

3 – Note 23

« Appendix I : Summary of the Points of Agreement Reached in the Previous Nine Conferences on Cybernetics », par Warren McCulloch, *in Macy* 10 Ce résumé fut distribué à l'avance aux participants de la dixième conférence. *Macy* 10 contient également un index cumulatif pour les Actes publiés (1949-1953).

— Retour au texte —

3 – Note 24

Plus particulièrement, les archives de McCulloch détenues par l'American Philosophical Society, à Philadelphie.

— Retour au texte —

3 – Note 25

S. HEIMS, « Encounter of Behavioral Sciences with new machine-organism analogies in the 1940's », *loc. cit. ;* et *The Cybernetics Group, op. cit.,* p. 235-238.

— Retour au texte —

3 – Note 26

Source : « A Note by the Editors », *in Macy* 8, p. XX.

— Retour au texte —

3 – Note 27

Voir S. HEIMS, « Encounter… », *loc. cit.,* p. 369.

— Retour au texte —

3 – Note 28

S. Heims, *John von Neumann and Norbert Wiener, op. cit.,* p. 206.

— Retour au texte —

3 – Note 29

Voir les remarques introductives de Frank FREMONT-SMITH, *in Macy* 7, p. 7-8.

— Retour au texte —

3 – Note 30

A. KOESTLER et J. R. SMYTHIES (eds.), *Beyond Reductionism,* Hutchinson, 1969.

— Retour au texte —

3 – Note 31

Jean Piaget, invité mais absent, était représenté par Bärbel Inhelder.

— Retour au texte —

3 – Note 32

Ce qui ne va pas sans entraîner certaines injustices. En tout état de cause, nous avons systématiquement inclus les « membres » réguliers du groupe, n'opérant un certain tri que sur les « invités ». Le groupe des membres réguliers comprend au départ (entre parenthèses, année de naissance) : Gregory Bateson (1904), anthropologue ; Julian Bigelow (1913), ingénieur ; Gerhardt von Bonin (1890), neuro-anatomiste ; Lawrence Frank (1890), sciences sociales ; Frank Fremont-Smith (1895), directeur médical de la fondation Macy ; Ralph Gerard (1900), neurophysiologiste ; Molly Harrower (1906), psychologue (démissionnne après la 5e conférence) ; George Hutchinson (1903), écologiste ; Heinrich Klüver (1897), psychologue ; Lawrence Kubie (1896), psychanalyste ; Paul Lazarsfeld (1901), sociologue (disparaît après la 6e conférence) ; Kurt Lewin (1890), psychologue social (meurt juste avant la 3e conférence) ; Rafael Lorente de No'(1902), neurophysiologiste ; Warren McCulloch (1899), neuropsychiatre ; Margaret Mead (1901), anthropologue ; John von Neumann (1903), mathématicien ; Filmer Northrop (1893), philosophe ; Walter Pitts (1923), mathématicien ; Arturo Rosenblueth (1900), physiologiste ; Leonard Savage (1917), mathématicien ; et Norbert Wiener (1894), mathématicien. Il faut adjoindre à cette liste les membres qui rejoignirent le groupe plus tard (entre parenthèses, après l'année de naissance, indication de la première conférence à laquelle le membre a participé) : Alex Bavelas (1913 ; 5e), psychologue social ; Henry Brosin (1904 ; 2e), psychiatre ; Heinz von Foerster (1911 ; 6regory Bateson (1904), anthropologue ; Julian Bigelow (1913), ingénieur ; Gerhardt von Bonin (1890), neuro-anatomiste ; Lawrence Frank (1890), sciences sociales ; Frank Fremont-Smith (1895), directeur médical de la fondation Macy ; Ralph Gerard (1900), neurophysiologiste ; Molly Harrower (1906), psychologue (démissionnne après la 5e conférence) ; George Hutchinson (1903), écologiste ; Heinrich Klüver (1897), psychologue ; Lawrence Kubie (1896), psychanalyste ; Paul Lazarsfeld (1901), sociologue (disparaît après la 6e, ingénieur électricien ; Donald Marquis (1908 ; 2e), psychologue ; Theodore Schneirla (1902 ; 2e), psychologie comparée ; Hans Lukas Teuber (1916 ; 4e), psychologue. Parmi les invités, on note (entre parenthèses : numéro de la ou des conférences auxquelles l'intéressé a participé) : Harold Abramson (6) ; W. Ross Ashby (9) ; Yehoshua Bar-Hillel (10) ; Herbert Birch (8) ; Eilhardt von Domarus

(5) ; Max Delbrück (5), Erik Erikson (3) ; Leon Festinger (3) ; Roman Jakobson (5) ; Clyde Kluckhohn (3, 4) ; Wolfgang Köhler (4) ; Joseph Licklider (7) ; William Livingston (2) ; Donald MacKay (8) ; Henry Quastler (9, 10) ; Claude Shannon (7, 8, 10) ; John Stroud (6, 7) ; W. Grey Walter (10) ; Heinz Werner (7) ; Jerome Wiesner (9), John Z. Young (9). Source : S. HEIMS, *The Cybernetics Group, op. cit.,* p. 285-286.

— Retour au texte —

3 – Note 33

L'expression est de Heinz VON FOERSTER dans « Circular Causality » *in McC. Intro.*

— Retour au texte —

3 – Note 34

Correspondance entre von Neumann et Wiener de novembre 1946, citée par S. HEIMS, *John von Neumann and Norbert Wiener, op. cit.,* p. 204.

— Retour au texte —

3 – Note 35

Lettre de Delbrück à S. HEIMS, *ibid.,* p. 205 et 475.

— Retour au texte —

3 – Note 36

L'œuvre de Henri ATLAN a non seulement jeté un éclairage complet sur cette histoire intellectuelle, elle l'a prolongée dans des directions neuves. Voir, entre autres, *L'Organisation biologique et la Théorie de l'information*, Hermann, Paris, 1972 (nouvelle éd., 1990) ; et *Entre le cristal et la fumée. Essai sur l'organisation du vivant*, Seuil, Paris, 1979.

— Retour au texte —

3 – Note 37

Macy 6, p. 150-157.

— Retour au texte —

3 – Note 38

Heinz VON FOERSTER, « Circular Causality », *in McC. Intro*, p. 8-9.

— Retour au texte —

3 – Note 39

Le seul recueil existant des écrits de VON FOERSTER s'intitule *Observing Systems* (titre qui, il est vrai, peut s'entendre de deux façons : « Systèmes observants » et « Observation des systèmes ») ; Intersystems Publications, 1981 (F. VARELA, éd.).

— Retour au texte —

3 – Note 40

Cité par S. HEIMS dans « Gregory Bateson and the Mathematicians », *loc. cit.,* p. 141.

— Retour au texte —

3 – Note 41

Ibid., p. 146-147 et 149-151.

— Retour au texte —

3 – Note 42

Steve Heims, communication orale.

— Retour au texte —

3 – Note 43

Cf. *Macy* 6, p. 9-10 ; *Macy* 7, p. 7-8 ; *Macy* 8, p. VII-VIII.

— Retour au texte —

3 – Note 44

Tel. Mech., p. 194. Cette citation de Haldane est en fait enchâssée dans une citation de Hopkins.

— Retour au texte —

3 – Note 45

Cité par S. HEIMS, *The Cyhernetics Group, op. cit.,* p. 173-174.

— Retour au texte —

3 – Note 46

Ibid., p. 69-71 et tout le chapitre 7, « The Macy Foundation and Worldwide Mental Health ». Le livre de Steve Heims est pour l'essentiel consacré à l'analyse de ces rapports entre sciences et société à l'époque de la cybernétique naissante.

— Retour au texte —

3 – Note 47

Tel. Mech.

— Retour au texte —

3 – Note 48

Macy 6, p. 151.

— Retour au texte —

3 – Note 49

Voir les remarques introductives de Fremont-Smith à la huitième conférence Macy (*Macy* 8, p. VIII) ; et aussi la discussion qui suit l'exposé de Kubie à la septième conférence (*Macy* 7, p. 224 et 234-235).

— Retour au texte —

3 – Note 50

Voir les interventions de Mead en *Macy* 7, p. 234, et *Macy* 8, p. 106.

— Retour au texte —

3 – Note 51

Cf. S. HEIMS, « Gregory Bateson and the Mathematicians », *loc. cit.*, p. 143. Boring ne participait pas aux conférences Macy.

— Retour au texte —

3 – Note 52

Source : MCCULLOCH, « An Account of the first three Conferences on Teleological Mechanisms ». Cf. note 22 ci-dessus.

— Retour au texte —

3 – Note 53

W. PITTS et W. MCCULLOCH, « How we know Universals : The Perception of Auditory and Visual Forms », 1947 ; repris dans *Embodiments of Mind*, *op. cit.*

— Retour au texte —

3 – Note 54

Macy 8, p. 104 et 131. Toute la discussion qui suit l'exposé de Kubie : « Communication between sane and insane : Hypnosis » est une des plus chaudes de toutes les conférences Macy.

— Retour au texte —

3 – Note 55

Macy 6, p. 9 et 101-107.

— Retour au texte —

3 – Note 56

Macy 8, p. 128 et 132.

— Retour au texte —

3 – Note 57

Ibid., p. 128-129.

— Retour au texte —

3 – Note 58

Ibid., p. 132.

— Retour au texte —

3 – Note 59

Macy 6, p. 148.

— Retour au texte —

3 – Note 60

Macy 9, p. 149.

— Retour au texte —

3 – Note 61

Hixon, p. 112.

— Retour au texte —

3 – Note 62

Macy 7, p. 11 et 47.

— Retour au texte —

3 – Note 63

McCulloch, « Why the Mind is in the Head », *in Hixon*, p. 43 ; repris dans *Embodiments of Mind, op. cit.,* p. 73.

— Retour au texte —

3 – Note 64

Dans « Circular Causality » (*in MCC. Intro, loc. cit.*). von Foerster cite le même passage de la communication de MCCulloch, et l'interprète, non pas comme un plaidoyer fait au nom de la physique, mais comme, déjà, l'annonce de la transformation de la cybernétique en épistémologie.

— Retour au texte —

3 – Note 65

« A Note by the Editors », *in Macy* 8, p. XI-XIII.

— Retour au texte —

3 – Note 66

Macy 7, p. 49.

— Retour au texte —

3 – Note 67

Cf., par exemple, le comportement de Bateson lors de la discussion qui suit la communication de Gerard sur les synapses inhibitrices et excitatrices, à la neuvième conférence (*Macy* 9, p. 137-138, 140, 146).

— Retour au texte —

3 – Note 68

Voir en *Macy* 7 les interventions de Gerard (p. 11-12), de Wiener (p. 18) et de Stroud (p. 28).

— Retour au texte —

3 – Note 69

Voir la sortie de Bigelow en *Macy* 9, p. 67-68 et 70.

— Retour au texte —

4 – Note 1

Voir, par exemple, les pages consacrées à ce tournant cognitif par l'un des meilleurs connaisseurs français de la philosophie de l'esprit, Pascal ENGEL, dans son livre : *États d'Esprit*, Alinéa, Aix-en-Provence, 1992 (p. 7-10) ; nouvelle édition. *Introduction à la philosophie de l'esprit*, La Découverte, Paris, 1994.

— Retour au texte —

4 – Note 2

Joëlle Proust, « L'intelligence artificielle comme philosophie », *Le Début*, n° 47, novembre-décembre 1987.

— Retour au texte —

4 – Note 3

Op. cit., p. XVII.

— Retour au texte —

4 – Note 4

J. MOSCONI, *La Constitution de ta théorie des automates, op. cit.*, t. I. p. 138.

— Retour au texte —

4 – Note 5

D. ANDLER, « From paleo to neo-connectionism », *in* G. VAN DER VIJVER (ed.). *New Perspectives on Cybernetics,* Kluwer, Dordrecht, 1992, p. 125-146.

— Retour au texte —

4 – Note 6

Cf. David E. RUMELHART, James L. MCCLELLAND & the PDP RESEARCH GROUP, *Parallel Distributed Processing : Explorations in the Mirostructure of Cognition*, t. I : *Foundations*, MIT Press, 1986.

— Retour au texte —

4 – Note 7

J. PROUST, *loc. cit.,* p. 98-99.

— Retour au texte —

4 – Note 8

On distingue aussi les sensations et, depuis une quinzaine d'années, les états de « connaissance tacite », ou états « sub-doxastiques ».

— Retour au texte —

4 – Note 9

R.M. CHISHOLM, *Perceiving* : *A Philosophical Study,* Ithaca, Cornell, 1957. Voir aussi un texte antérieur : « Sentences about believing », *Proceedings of the Aristotelian Society* 56 (1956).

— Retour au texte —

4 – Note 10

Ibid., p. 298.

— Retour au texte —

4 – Note 11

Pour une vue d'ensemble de ce problème et des diverses positions théoriques qu'il a engendrées, on étudiera avec profit le livre de Pascal Engel, *États d'Esprit, op. cit*.

— Retour au texte —

4 – Note 12

Repris dans D. DAVIDSON, *Essays on Actions and Events,* Oxford University Press, 1980 ; trad. fr. par Pascal ENGEL, *Actions et Événements*, PUF, Paris, 1994.

— Retour au texte —

4 – Note 13

P. ENGEL, *Étais d'Eprit, op. cit.,* p. 11-12.

— Retour au texte —

4 – Note 14

J. SEARLE, « Minds, Brains, and Programs », *The Behavioral and Brain Sciences*, t. III, 1980.

— Retour au texte —

4 – Note 15

W.V.O. QUINE, article « Mind versus Body », *Quiddities : An Intermittently Philosophical Dictionary*, Belknap Press of Harvard University Press, 1987.

— Retour au texte —

4 – Note 16

F. BRENTANO, *Psychologie vom empirischen Standpunkt*, Vienne, 1874 ; il existe une traduction française due à Maurice de Gandillac, parue chez Aubier en 1944.

— Retour au texte —

4 – Note 17

Ma traduction. Sur le rapport de Husserl à Brentano, surtout en tant qu'il se situe en tronc commun par rapport à la bifurcation qui mène d'un côté à Heidegger et de l'autre à Sartre, on consultera avec profit Alain RENAUT, *Sartre, le dernier philosophe,* Grasset, Paris, 1993 ; plus spécialement p. 88-102.

— Retour au texte —

4 – Note 18

F. BRENTANO, *Psychologie*, *op. cit. ;* ma traduction.

— Retour au texte —

4 – Note 19

A. RENAUT, *Sartre*, *op. cit.,* p. 93 et 97-98.

— Retour au texte —

4 – Note 20

Voir Linda MCALISTER, « Chisholm and Brentano on Intentionality », *Review of Metaphvsics,* XXVIII, n° 2, 1974, p. 328-338 ; et, surtout, le travail remarquable réalisé actuellement par Stefano Franchi au département de philosophie de l'université Stanford. L'affaire est, en vérité, plus compliquée car Brentano lui-même en viendra plus tard, dès 1905 puis dans la seconde édition de *Psychologie*, publiée en 1911, à rejeter la thèse selon laquelle les « phénomènes physiques » sont contenus à l'intérieur des phénomènes mentaux (cette étrange *Kehre* amenant Brentano à adopter ce qu'il nomme lui-même un « réisme »). Le problème est que Chisholm ne se réfère jamais qu'à l'édition de 1874. Certains philosophes analytiques spécialistes de Brentano soutiennent qu'il n'y a jamais eu de retournement et que Brentano était un « réiste » dès le départ. C'est donc la tradition husserlienne qui se serait méprise gravement ! Cf. Richard AQUILA, *Intentionality : A Stuclv of Mental Aets,* The Pennsylvania State University Press, 1977.

— Retour au texte —

4 – Note 21

§45 : « The Double Standard », *Word and Object*, MIT Press, 1960.

— Retour au texte —

4 – Note 22

W.V.O. QUINE, *Word and Object,* op. cit., p. 221.

— Retour au texte —

4 – Note 23

J. -P. Sartre, *L'existentialisme est un humanisme*, Nagel, Paris, 1970, p. 17 (origin. 1946). Voir cependant la réhabilitation partielle proposée par A. RENAUT dans son *Sartre*, *op. cit.*

— Retour au texte —

4 – Note 24

On trouvera une introduction accessible à ces travaux dans F. FOGELMAN SOULIÉ (ed), *Les Théories de la complexité. Autour de l'œuvre d'Henri Atlan*, Seuil, Paris, 1991.

— Retour au texte —

4 – Note 25

S. KAUFFMAN, *Origins of Order and Self Organisation in Evolution*, Oxford University Press, New York, 1991.

— Retour au texte —

4 – Note 26

F. VARELA, *Principles of Biological Autonomy*, North Holland, 1979 (trad, fr. : Autonomie *et Connaissance, op. cit.*).

— Retour au texte —

4 – Note 27

J. HOPFIELD, « Neural Networks and Physical Systems with Emergent Collective Computational Abilities » (1982), repris *in* J. ANDERSON et E. ROSENFELD (eds), *Neurocomputing*, MIT Press, 1988.

— Retour au texte —

4 – Note 28

D. AMIT, *Modeling Brain Function. The World of Attractor Neural Networks*, Cambridge University Press, 1989.

— Retour au texte —

4 – Note 29

J. PETITOT, « Morphodynamics and Attractor Syntax. Dynamical and morphological models for constituency in visual perception and cognitive grammar », *in* T. VAN GELDER et R. PORT (eds). *Mind as Motion*, Cambridge, MIT Press, 1994 ; et *La Philosophie transcendantale et le problème de l'objectivité*, Entretiens du Centre Sèvres (père F. MARTY éd.), Éditions Osiris, Paris, 1991.

— Retour au texte —

4 – Note 30

H. ATLAN, « Intentionality in Nature ». *Journal for the Theory of Social Behaviour*, 24 : 1, 1994, p. 67-87.

— Retour au texte —

4 – Note 31

J. PETITOT, « Phénoménologie naturalisée et morphodynamique : la fonction cognitive du synthétique *a priori* », *Philosophie et sciences cognitives* (J.-M. SALANSKIS éd.), *Intellectica*, 1992/3, 17, p. 79-126 ; et *Physique du sens*, Éd. du CNRS, Paris, 1992.

— Retour au texte —

4 – Note 32

F. VARELA, E. THOMPSON et E. ROSCH, The Embodied Mind, MIT Press, 1991 (trad. fr. par V. Havelange, *L'Inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, Seuil, Paris, 1993).

— Retour au texte —

4 – Note 33

Pierre LÉVY a consacré une étude à l'influence de Wittgenstein sur la cybernétique. Voir son « Wittgenstein et la cybernétique », *Cahiers du CRÉA*, n° 7, novembre 1985.

— Retour au texte —

4 – Note 34

Steve HEIMS a pu rétablir les éléments essentiels de cette rencontre qui, rappelons-le, n’a pas donné lieu à compte rendu écrit. Voir le chapitre 10 de *The Cybernetics Group*, *op. cit.*

— Retour au texte —

4 – Note 35

Original, 1959 ; repris dans *Embodiments of Mind*, *op. cit*p. 230-255.

— Retour au texte —

4 – Note 36

Paul RICŒUR, « L’herméneutique et le structuralisme », *Espriu* novembre 1963, p. 618.

— Retour au texte —

4 – Note 37

Cl. LÉVI-STRAUSS, « Introduction à l'œuvre de Marcel Mauss », *in* M. MAUSS, *Sociologie et Anthropologie*, PUF, Paris, 1973.

— Retour au texte —

4 – Note 38

Trad. fr. dans M. HEIDEGGER, *Essais et Conférences*, Gallimard, Paris, 1958, p. 44.

— Retour au texte —

4 – Note 39

J. LACAN, *Le Séminaire*, livre II, Seuil, Paris, 1978, chap. XI et XVI, « Séminaire sur *La Lettre volée* ».

— Retour au texte —

4 – Note 40

Pour un bilan de ces travaux, on pourra consulter J. -P. DUPUY, « Self-Reference in Literature », *Poetics*, n° 18, 1989, p. 491-515.

— Retour au texte —

4 – Note 41

Cf. *Le Séminaire*, livre II, *op. cit.,* p. 111-112.

— Retour au texte —

4 – Note 42

Ibid., p. 351.

— Retour au texte —

4 – Note 43

M. Heidegger, *Essais et Conférences, op. cit.,* p. 44-45.

— Retour au texte —

4 – Note 44

Ph. BRETON, « La cybernétique et les ingénieurs dans les années cinquante », *Culture Technique*, n° 12, mars 1984, p. 160. Voir aussi « Pourquoi les machines analogiques ont-elles disparu ? » *Milieux*, juin 1985 ; et « Quelques problèmes posés par l'émergence d'un nouveau type d'ingénieurs : les ingénieurs de la connaissance », *Cahiers du CEFI,* juin 1985. La grande différence entre les origines américaines de la cybernétique et son acclimatation en France, c'est que celle-ci s'est faite principalement dans et par le milieu des ingénieurs, alors que les préoccupations de ce milieu n'ont joué qu'un rôle très partiel dans le développement de la cybernétique sur son sol natal. On consultera avec profit les ouvrages de Ph. BRETON (*L'Utopie de la communication*, La Découverte, Paris, 1992 ; *L'Explosion de la communication*, en coll. avec S. PROULX, La Découverte, Paris, 1994).

— Retour au texte —

4 – Note 45

Flammarion, Paris, 1980, p. 332-333.

— Retour au texte —

4 – Note 46

G. HOTTOIS, *Le Signe et la Technique*, Aubier, Paris, 1984, p. 150.

— Retour au texte —

4 – Note 47

J. LACAN, *Le Séminaire*, livre II, *op. cit.,* p. 367.

— Retour au texte —

4 – Note 48

M. Jeannerod, *Le Cerveau-machine*, Fayard, Paris, 1983, p. 156. Souligné par l’auteur.

— Retour au texte —

4 – Note 49

V. DESCOMBES, *Le Même et l'Autre.* Minuit, Paris, 1979, p. 123-124.

— Retour au texte —

4 – Note 50

Cf. *Cybernetics, op. cit.,* p. 13.

— Retour au texte —

4 – Note 51

Sur l'effet de choc produit sur Wiener (et. comparativement, sur von Neumann) par la guerre, voir le chapitre 9, « The Watershed » de *John von Neumann and Norbert Wiener*, *op. cit.* p. 157 et 343.

— Retour au texte —

4 – Note 52

Sur la façon dont le style de pensée parfois relâché de Wiener était reçu par ses collègues, voir S. HEIMS, *John von Neumann and Norbert Wiener*, *op. cit.,* p. 157 et 343.

— Retour au texte —

4 – Note 53

A l'exception bien sûr du physiologiste « cybernéticien » Rosenblueth ; mais aussi, ce qui est plus notable, de Lorente de No'. Voir l'admiration qu'il manifeste à l'égard de McCulloch, à la suite de l'exposé de celui-ci au symposium Hixon (*Hixon,* p. 57-58). Voir en revanche les réserves exprimées par le grand neurophysiologiste Ralph Gerard, pourtant collaborateur de McCulloch et Pitts, et pilier des conférences Macy. (Interview avec S. HEIMS, *in John von Neumann and Norbert Wiener, op. cit.,* p. 359.)

— Retour au texte —

4 – Note 54

Sur l'effet de choc produit sur Wiener (et. comparativement, sur von Neumann) par la guerre, voir le chapitre 9, « The Watershed » de *John von Neumann and Norbert Wiener*, *op. cit.*

— Retour au texte —

4 – Note 55

Voir les chapitres 13 : « Wiener, the Independent Intellectual : Technology as Applied Moral and Social Philosophy », et 15 : « Wiener's later years : Again the Golem », du même ouvrage ; et : Philippe BRETON, « Présupposés anthropologiques et pensée du social chez les ingénieurs : un facteur sous-estimé du développement des technologies de l'information de l'après-guerre ». *Cahiers STS,* CNRS, 1986.

— Retour au texte —

4 – Note 56

Une histoire de la cybernétique centrée sur la tension entre les styles de pensée propres à Wiener et à McCulloch donnerait une vision des choses sensiblement différente de celle qui ressort du livre de Steve Heims, lequel oppose Wiener et von Neumann.

— Retour au texte —

5 – Note 1

Macy 6, p. 150. Alors que les noms communs sont généralement neutres en anglais, *ship* (= bateau) a une connotation féminine, et l'on dit d'une *ship :* « *she* ».

— Retour au texte —

5 – Note 2

Tel. Mech., p. 203 ; Wiener reprendra la même idée au cours de la troisième conférence Macy, en 1947 (Cf. Me CULLOCH, « An Account of the first three conferences on teleological mechanisms »).

— Retour au texte —

5 – Note 3

Cf. par exemple *Macy* 6, p. 107.

— Retour au texte —

5 – Note 4

Tel. Meclu p. 207-208.

— Retour au texte —

5 – Note 5

Macy 7, p. 26.

— Retour au texte —

5 – Note 6

Ibid., p. 20-21.

— Retour au texte —

5 – Note 7

Ibid., p. 43-44.

— Retour au texte —

5 – Note 8

Ibid., p. 45-46.

— Retour au texte —

5 – Note 9

Wiener utilise les expressions *region of attraction* (*Macy* 6, p. 201) et *field of attraction* (*Macy* 7, p. 21).

— Retour au texte —

5 – Note 10

Macy 6, p. 202. Wiener développe ces idées à la sixième conférence, en l'absence de von Neumann. A la septième, von Neumann les introduit à son tour (*Macy* 7, p. 19-20) avec une très grande clarté, et Wiener lui emboîte le pas, affirmant qu'il ne fait que « commenter ce que le professeur von Neumann vient de dire » (p. 21).

— Retour au texte —

5 – Note 11

Macy 7, p. 18-19.

— Retour au texte —

5 – Note 12

J.-P. CHANGEUX, *L'Homme neuronal*, *op. cit.,* p. 137 et p. 116.

— Retour au texte —

5 – Note 13

Macy 6, p. 88-89. Le rôle de ces messages humoraux est également évoqué dans *Cybernetics, op. cit.,* p. 129.

— Retour au texte —

5 – Note 14

De même, mais dans un tout autre style, que von Neumann l'indépendant. Les deux coauteurs de Wiener pour l'article de 1943 étaient sans doute les plus rigides et dogmatiques. Voir la sortie de Bigelow à rencontre de Gerard, à l'issue de la discussion que nous venons de commenter, sur les rôles respectifs du numérique et de l'analogique dans le système nerveux central, *in Macy* 7, p. 47.

— Retour au texte —

5 – Note 15

Source : W. McCulloch, « An Account of the first three conferences on teleological mechanisms », *loc. cit.* Wiener est ici le professeur.

— Retour au texte —

5 – Note 16

McCulloch, *Macy* 6, p. 97.

— Retour au texte —

5 – Note 17

Le professeur est ici Licklider, qui se définit comme occupant une position intermédiaire entre celles du psychologue et du mathématicien. *Macy* 7, p. 228-229.

— Retour au texte —

5 – Note 18

Macy 7, p. 58 *sq :* à noter la discussion avec Teuber qui préfère parler d'« équivalence de stimuli » plutôt que de « redondance de l'information » (p. 95-96).

— Retour au texte —

5 – Note 19

Ibid., p. 66.

— Retour au texte —

5 – Note 20

Cl. SHANNON, « The Redundancy of English », *Macy* 7, p. 123 *sq.*

— Retour au texte —

5 – Note 21

Voir *supra,* chap. 2, p. 55.

— Retour au texte —

5 – Note 22

Hixon, p. 55.

— Retour au texte —

5 – Note 23

Cf. J. -P. CHANGEUX, *L'Homme neuronal, op. cit.,* chap, VII : « Épigenèse », spécialement p. 273 et 302.

— Retour au texte —

5 – Note 24

Henry Quastler, *Macy* 9, spécialement p. 173 et 179-180. On trouvera un exposé plus complet des calculs de Quastler dans Henri Atlan, *L'Organisation biologique et la Théorie de l'information*, Hermann, Paris, 1972, chap. 5.

— Retour au texte —

5 – Note 25

Raymond RUYER, *La Cybernétique et l'Origine de l'information*, Flammarion, Paris, 1954, p. 25-26.

— Retour au texte —

5 – Note 26

Ibid, p. 13.

— Retour au texte —

5 – Note 27

F. DRETSKE, *Knowledge and the Flow of Information*, Blackwell, 1981 ; et *Explaining Behavior – Reasons in a World of Causes*, The MIT Press, 1988.

— Retour au texte —

5 – Note 28

McCulloch, *Hixon*, p. 46.

— Retour au texte —

5 – Note 29

Macy 7, p. 154 et 147.

— Retour au texte —

5 – Note 30

Voir l'échange entre Shannon, Licklider, Stroud, Pitts et Savage *in Macy* 7, p. 154-155.

— Retour au texte —

5 – Note 31

Ibid., p. 140.

— Retour au texte —

5 – Note 32

Ibid., p. 141-146.

— Retour au texte —

5 – Note 33

H. VON FOERSTER, « On Self-Organizing Systems and their Environments », 1960, repris *in Observing Systems,* Intersystems Publications, 1981.

— Retour au texte —

5 – Note 34

H. ATLAN, *Entre le cristal et la fumée*, Seuil, Paris, 1979.

— Retour au texte —

5 – Note 35

Macy 8, p. XIII.

— Retour au texte —

5 – Note 36

N. WIENER, *I Am a Mathematiciain* MIT Press, 1964. Cité par S. HEIMS, *John von Neumann and Norbert Wiener*, *op. cit.,* p. 155-156 (chapitre 7, significativement intitulé : « The Foundation : Chaos or Logic ? »).

— Retour au texte —

5 – Note 37

Macy 7, p. 153-154.

— Retour au texte —

5 – Note 38

G. SIMONDON, *Du mode d'existence des objets techniques*, Aubier-Montaigne, Paris, 1969, p. 49. Référence aux Actes des conférences Macy, p. 263.

— Retour au texte —

5 – Note 39

Ibid., p. 137.

— Retour au texte —

5 – Note 40

Voir le débat entre MacKay et Savage *in Macy* 8, p. 204-206. Savage, qui défendait une conception subjectiviste des probabilités, ne croyait pas que l'incertain pût être toujours réductible au probabilisable.

— Retour au texte —

5 – Note 41

Il est vrai que Wiener, le plus sensible à la dialectique entre ordre et désordre, disparaît dès la huitième conférence.

— Retour au texte —

5 – Note 42

Carnap, qui avait été le professeur de Pitts à Chicago, fut invité à participer, mais dut refuser pour raisons de santé.

— Retour au texte —

5 – Note 43

Macy 6, p. 138. Voir le débat entre Stroud, Klüver, Kubie, Teuber, McCulloch et Wiener, *ibid.,* p. 57-59.

— Retour au texte —

5 – Note 44

Macy 8, p. 217-218.

— Retour au texte —

5 – Note 45

Cf. McCULLOCH, « A Heterarchy of Values Determined by the Topology of Nervous Nets », repris dans *Embodiments of Mind, op. cit.*

— Retour au texte —

5 – Note 46

Cf. S. HEIMS, *John Von Neumann and Norbert Wiener, op. cit.,* p. 306-307 ; et *The Cybernetics Group, op. cit*., p. 110.

— Retour au texte —

5 – Note 47

Ibid., chap. 9 et 10.

— Retour au texte —

5 – Note 48

L'un et l'autre chez Harvard University Press.

— Retour au texte —

5 – Note 49

Cf. J.-P. DUPUY, *Introduction aux sciences sociales. Logique des phénomènes collectifs,* Ellipses, Paris, 1992.

— Retour au texte —

5 – Note 50

Macy 8, p. 42.

— Retour au texte —

5 – Note 51

R. RUYER, *op. cit.,* p. 62-63.

— Retour au texte —

5 – Note 52

J.-C. BEAUNE, *op. cit.* p. 332 et 328.

— Retour au texte —

5 – Note 53

G. SIMONDON, *op. cit., p. 104-105.*

— Retour au texte —

5 – Note 54

Ibid., p. 103.

— Retour au texte —

5 – Note 55

Cf. Georges CANGUJLHEM, « Le tout et la partie dans la pensée biologique » (1966), repris dans *Études d'histoire et de philosophie des sciences.*

— Retour au texte —

5 – Note 56

Voir, par exemple, la discussion qui suit l'exposé de Shannon sur sa machine à résoudre des problèmes de labyrinthe *in Macy* 8, p. 173-180, et aussi cette réflexion de Bateson lors de la discussion sur l'homéostat d'Ashby *in Macy* 9, p. 106 : Bateson demande à l'écologiste Hutchinson si l'on peut dire qu'un écosystème « apprend » selon la même logique que l'homéostat (dont Ashby vient de prétendre qu'il *était* un cerveau). A quoi Hutchinson répond : oui, absolument !

— Retour au texte —

5 – Note 57

Cf. P. S. CHURCHLAND, *Neurophilosophy,* MIT, Bradford Books, 1986 ; et P. M. CHURCHLAND, « Eliminative materialism and propositional attitudes », *Journal of philosophy,* 1979.

— Retour au texte —

5 – Note 58

J.-P. CHANGEUX, *op. cit.*, p. 211.

— Retour au texte —

5 – Note 59

Cf. F. VARELA, *Principles of Biological Autonomy, op. cit*.

— Retour au texte —

5 – Note 60

Travaillant ensemble à la recherche dont ce livre est issu, et dont je rappelle les circonstances dans l'avant-propos, Isabelle Stengers et moi sommes tombés, indépendamment l'un de l'autre, sur la figure, aujourd'hui trop peu connue, de Paul Weiss. Cette double rencontre, non planifiée, nous a convaincus, elle et moi, du rôle important qu'avait joué Weiss dans toute cette histoire. Voir I. STENGERS, « Les généalogies de l'auto-organisation », *Cahiers du CRÉA*, n° 8, novembre 1985.

— Retour au texte —

5 – Note 61

N. BOHR, *Light and Life*, 1932 ; trad. fr. *Physique atomique et connaissance humaine*, Gauthiers-Villars, Paris, 1972.

— Retour au texte —

5 – Note 62

Cf. I. STENGERS, *loc. cit.*, p. 64.

— Retour au texte —

5 – Note 63

I. Stengers, qui développe avec brio tous ces points dans le travail cité ci-dessus, estime que cette solidarité entre l'auto-organisation des embryologistes et celle de l'École de Bruxelles ne s'étend pas à la théorie des systèmes autopoiétiques développée par l'École chilienne (Maturana et Varela), ni à la théorie de l'auto-organisation mise au point par Henri Atlan. Ces deux dernières théories resteraient prisonnières des schémas mécanistes de la seconde cybernétique. J'ai tendance à minimiser l'opposition, mais ce n'est pas le lieu d'en discuter ici. Cf. P. DUMOUCHEL et J. P. DUPUY (éds), *L'Auto-organisation, de la physique au politique*, Seuil, Paris, 1983 (Actes du colloque de Cerisy de juin 1981).

— Retour au texte —

5 – Note 64

P. WEISS, « Le concept fondamental de système hiérarchique », trad. fr. dans *L'Archipel scientifique,* Maloine, Paris, 1974, p. 116-117 ; cité par I. Stengers, *loc. cit.*

— Retour au texte —

5 – Note 65

A. KOESTLER et J. R. SMYTHIES (eds), *Beyond Reductionisnu* Hutchinson, 1969, p. 62. Von Bertalanffy distingue comme autres approches systémiques, à côté de la cybernétique, la théorie de l'information, la théorie des automates, la théorie des jeux et la théorie des files d'attente – comme si celles-ci, pour les trois premières en tout cas, n'avaient pas été parties intégrantes de la cybernétique !

— Retour au texte —

5 – Note 66

L. VON BERTALANFFY, « The Meaning of General System Theory » *in General System Theory*, Braziller, New York, 1980, p. 44 ; cité par I. STENGERS. L'« équipotentialité », ou encore « équifinaiité », est définie par Bertalanffy comme la propriété de certains processus qui aboutissent « au même but, quel que soit le point de départ et quelles que soient les voies suivies pour atteindre ce but ». C'est ainsi que les fameuses expériences de Hans Driesch sur l'embryon d'oursin ont montré, dès 1891, que l'on obtient toujours une larve complète, que l'on parte d'un germe complet, d'un demi-germe ou d'un quart de germe.

— Retour au texte —

5 – Note 67

Bcvond Recluctionism, op. cit., p. 19.

— Retour au texte —

5 – Note 68

Ibid, p. 15, n. 3.

— Retour au texte —

5 – Note 69

Ibid., p. 10-11.

— Retour au texte —

5 – Note 70

P. WEISS, « Le déterminisme stratifié des systèmes vivants », 1969, repris *in L'Archipel scientifique, op. cit.,* p. 205.

— Retour au texte —

5 – Note 71

On observe avec amusement que dans sa « Note by the Editors », publiée au début de *Macy* 8, p. XVI (1951), Heinz VON FOERSTER « récupère » l'argument de Lashley et l'attribue à la sagacité des cybernéticiens alliée aux connaissances… de Lashley, dans leur lutte commune contre la vision simpliste du système nerveux central comme organe réflexe, *input-output.* Ce qui lui permet ce tour de prestidigitation, c'est évidemment l'attention portée par McCulloch, Kubie et Lorente de No'aux circuits réverbérants. Pour sa part, von Foerster se fera plus tard le champion de cette conception du système nerveux central comme système « autonome », créateur d'un monde. Cf. « On Constructing a Reality », 1973, repris *in Observing Systems, op. cit.*

— Retour au texte —

5 – Note 72

Lashley, *Hixon*, p. 70-71 et 112-113.

— Retour au texte —

5 – Note 73

J.-P. CHANGEUX, *L'Homme neuronal, op. cit.,* p. 107. Changeux se réfère aux expériences de Lashley, mais en profite pour ranger celui-ci, à l'instar de Flourens, dans le camp des « spiritualistes ». (*ibid.,* p. 335-336). C'est sans doute trop simple, et tout ce qui diffère du monisme réductionniste et matérialiste à la McCulloch… ou à la Changeux, ne peut être rangé dans le même sac. Comment expliquer, sinon, que Lashley fasse au début de sa communication au symposium Hixon la déclaration que nous avons rapportée ci-dessus, p. 87 ? Autre chassé-croisé intéressant : le neurobiologiste Roger Sperry, qui fut à la fois l'élève de Paul Weiss et celui de Karl Lashley, devait par ses travaux – qui lui valurent le prix Nobel de médecine en 1981, avec Hubel et Wiesel – réfuter les théories de ses deux mentors. Les expériences qu'il a menées sur le triton tendent à montrer que le câblage du cerveau est très spécifique, et qu'à chaque fonction correspond un chemin neuronal particulier.

— Retour au texte —

5 – Note 74

P. Weiss, *Hixon*, p. 72-74 et 140-142.

— Retour au texte —

5 – Note 75

L’expression est de Köhler, *Hixon*, p. 65.

— Retour au texte —

5 – Note 76

P. WEISS, *Dynamics of Development : Experiments and Inferences,* Académie Press, New York, 1968. I. STENGERS remarque, dans ses « Généalogies de l'auto-organisation », loc. cit., que cette citation a plusieurs fois été faite par I. Prigogine, en particulier dans un exposé de 1969, où elle précède immédiatement ce qui a peut-être été la première mention, par l'école thermodynamique de Bruxelles, du terme « auto-organisation ».

— Retour au texte —

5 – Note 77

P. Weiss, *Bevond Reductionism, op. cit.,* p. 6-9 et 45-46.

— Retour au texte —

5 – Note 78

Macy 9, p. 11.

— Retour au texte —

5 – Note 79

Macy 8, p. 52-54 et 80-83.

— Retour au texte —

5 – Note 80

Macy S, p. 124-125.

— Retour au texte —

5 – Note 81

H. VON FOFRSTER, « Morality Play », *The Sciences*, octobre 1981, p. 24-25.

— Retour au texte —

5 – Note 82

G. SIMONDON, *Du mode d'existence des objets techniques, op. cit.*, p. 49.

— Retour au texte —

5 – Note 83

H. VON FOERSTER, « Note by the Editors », *Macy* 8, p. XVII. Jean-Claude BEAUNE, pour sa part, écrit : « Les ordinateurs sont des systèmes mathématiques et logiques mis en acte, des théories incarnées, contemporaines des expériences qu'elles suscitent [...] Tout modèle est, en droit, machine à raisonner. L'ordinateur propose de cette activité une réinterprétation spécifiquement technique et, si l'on veut, fonctionnaliste » (*L'Automate et ses mobiles, op. cit.,* p. 315).

— Retour au texte —

5 – Note 84

Jerome LETTVIN, « Warren and Walter », p. 11, in *McC. Intro.*

— Retour au texte —

5 – Note 85

A. ROSENBLUETH et N. WIENER, « The Role of Models in Science », *Philosophy of Science,* octobre 1945, vol. 12. n° 4.

— Retour au texte —

5 – Note 86

Cf. l'interprétation par Josiah ROYCE. de la définition de l'infini donnée par Dedekind en termes de « système auto-représentatif » *in* « The Concept of the Infinite », *Hilbert Journal* I, 1902, p. 21-45. Voir aussi le parti que Borges tire de cette figure dans « Magies partielles du *Quichotte* » *in Enquêtes*, Gallimard, Paris, 1957. Sur le rapport à la « mise en abyme » en général, voir J.-P. DUPUY, « Self-Reference in Literature », *Poetics*, 18, 1989, p. 491-515.

— Retour au texte —

5 – Note 87

Macy 9. p. 147.

— Retour au texte —

5 – Note 88

Macy 7. p. 19-51.

— Retour au texte —

5 – Note 89

Cf. P. LÉVY, « Wittgenstein et la cybernétique », *loc. cit.*

— Retour au texte —

5 – Note 90

Cf. Jacques DERRIDA, « La double séance », *in La Dissémination*, Seuil, Paris, 1972.

— Retour au texte —

5 – Note 91

Cité par F. VON HAYEK, *Individualism and Economic Order*, University of Chicago Press, 1976, p. 181-182.

— Retour au texte —

5 – Note 92

Cf., dans *Theory of Self-Reproducing Automata, op. cit.,* l'introduction de BURKS, p. 2-4, et la première conférence de VON NEUMANN, p. 33-35.

— Retour au texte —

5 – Note 93

Ihid., p. 47.

— Retour au texte —

5 – Note 94

Ibid., p. 53-56.

— Retour au texte —

5 – Note 95

D. DAVIDSON, « Mental Events » (1970), repris *in Essavs on Actions and Events*, Clarendon Press, Oxford, 1980, p. 214-215. Davidson précise que cette analogie ne saurait être poussée trop loin.

— Retour au texte —

5 – Note 96

Von Neumann, *Hixon,* p. 109-110.

— Retour au texte —

6 – Note 1

W. MCCULLOCH, « Finality and Form in Nervous Activity », conférence de 1946, publiée en 1952, reprise *in Embodiments of Mind, op. cit.* (voir p. 257 et 270). Rappelons que McCulloch et Pitts eurent *avant* von Neumann l'idée de considérer des réseaux aléatoires. Également remarquable est la réflexion par laquelle McCulloch conclut sa conférence, parce qu'elle annonce le débat d'aujourd'hui entre les néoconnexionnistes et les partisans de l'intelligence artificielle « orthodoxe ». Un réseau aléatoire peut apprendre, dit-il, et en particulier réaliser des inférences inductives, mais il faudrait de plus qu'il soit capable d'exprimer les règles *suivies* par lui, ce faisant. McCulloch pensait à l'époque, comme Daniel Dennett et beaucoup d'autres aujourd'hui, qu'il suffirait de quelques années pour résoudre ce problème (MCCULLOCH, *ibid,* p. 273-275 ; D. DENNETT, « The Logical Geography of Computational Approaches », *loc. cit.*).

— Retour au texte —

6 – Note 2

Pitts, *Macy* 9, p. 107-108.

— Retour au texte —

6 – Note 3

Hixon, p. 102.

— Retour au texte —

6 – Note 4

McCulloch, *Tel. Mech.*, p. 269.

— Retour au texte —

6 – Note 5

Que la biologie moléculaire fût obligée de recourir à des termes anthropomorphiques pour rendre compte, selon un schéma emprunté à la cybernétique, des propriétés intégratrices de l'organisme, était pour Paul Weiss, on l'a vu, le meilleur signe qui fût de l'impuissance constitutive de cette discipline (la cybernétique) à fournir des modèles du vivant. Voir sa communication au symposium *Bevond Reductionism. op. cit.,* p. 29 et 36.

— Retour au texte —

6 – Note 6

Birch, *Macy* 8, p. 154-157.

— Retour au texte —

6 – Note 7

Cf. par exemple la « Note by the Editors », *Macy* 8. spécialement p. XIV-XVII. Dans *La Cybernétique et l'Origine de l'information*, *op. cit.,* Raymond RUYER ne voit, à notre sens, que la moitié de la vérité lorsqu'il écrit : « On connaît les abus qui ont été faits de la notion de *feedback.* ou rétroaction, confondue avec un simple équilibrage physique et finissant par absorber dans la cybernétique tous les phénomènes physiques d'équilibre » (p. 234). L'absorption s'est faite aussi dans l'autre sens, privant de plus en plus le feedback de sa dimension d'organe de commande et de régulation. Raymond Ruyer ne peut voir cet aspect des choses parce qu'il tient, plus encore que la cybernétique, à l'information.

— Retour au texte —

6 – Note 8

Au colloque de Montpellier sur l'œuvre de Herbert Simon, organisé par Jean-Louis Le Moigne en janvier 1984, j'ai interrogé le fondateur de l'intelligence artificielle sur l'influence qu'avait exercée sur lui la cybernétique. « Le seul cybernéticien qui a eu un impact sur mon œuvre, répondit-il, est Ross Ashby. » Rappelons, pour aller à une autre extrémité du champ, que Jacques Lacan connaît le travail d'Ashby, et s'y réfère dans son Séminaire de l'année 1954-1955.

— Retour au texte —

6 – Note 9

Cf., par exemple, R. ASHBY, « Principies of the Self-Organizing System », 1962 ; repris *in Mechanisms of Intelligence : Ross Ashby's Writings on Cybernetics*, Intersystems Publications, 1981, p. 65-69.

— Retour au texte —

6 – Note 10

Toute la discussion *in Macy* 9, p. 74-108 et 151-154, est extrêmement savoureuse. De l'aveu de von Foerster, qui a transcrit les débats, les termes employés étaient en réalité beaucoup plus rudes que ceux qu'il a finalement retenus.

— Retour au texte —

6 – Note 11

Ibid., p. 73.

— Retour au texte —

6 – Note 12

I. STENGERS, « Les généalogies de l'auto-organisation », *loc. cit.,* p. 49 et 46.

— Retour au texte —

6 – Note 13

Selon l’expression de Zopf, collaborateur de von Foerster. au troisième des symposiums organisés par le Biological Computer Laboratory sur les systèmes à auto-organisation (1960).

— Retour au texte —

6 – Note 14

D. HOFSTADTER, *Gödei, Escher, Bach*, Basic Books, New York, 1979, p. 477-478 ; cité par N. Luhmann, « The Autopoiesis of Social Systems », *in* G. TEUBNER (ed.), *Autopoielic Law : A New Approach to Law and Society,* Walter de Gruyter, 1988

— Retour au texte —

6 – Note 15

R. DEBRAY, *Critique de la raison politique*, Gallimard, Paris, 1981, p. 256.

— Retour au texte —

6 – Note 16

I. STENGERS, « Les généalogies de l’auto-organisation », *loc. cit.,* p. 31.

— Retour au texte —

6 – Note 17

Les concepteurs de réseaux d'automates booléens ne cherchent pas à cacher la *surprise* que leur procure la découverte des propriétés remarquables de leurs machines. Cf. les communications de Françoise Fogelman-Soulié et Henri Atlan au colloque de Cerisy, *L'Auto-organisation*, *op. cit.*

— Retour au texte —

6 – Note 18

Voir *Beyond Reductionisnu op. cit*., p. 367-368 ; Stuart KAUFFMAN, « Behavior of Gene Nets », *in McC. Intro.*

— Retour au texte —

6 – Note 19

I. STENGKRS. « Les généalogies de l’auto-organisation », *loc. cit.,* p. 94.

— Retour au texte —

6 – Note 20

Ibid.. p. 99.

— Retour au texte —

6 – Note 21

Cf. *L'Homme neuronal, op. cit.*, p. 8-9.

— Retour au texte —

6 – Note 22

Cf. J.-P. DUPUY, « De l'émancipation de l'économie. Retour sur le "problème d'Adam Smith" », *L'Année sociologique*, 1987, 37.

— Retour au texte —

6 – Note 23

F. VON HAYEK, *Droit, législation et liberté*, vol. I, *Règles et ordre*, tr. fr. de R. Audoin, PUF, Paris, 1980, p. 21 et 35. Pour une présentation générale et une critique de la philosophie hayékienne de l'esprit et de son rapport à la philosophie sociale, voir J.-P. DUPUY, *Le Sacrifice et l'Envie*, Calmann-Lévy, Paris, 1992, chap. vin.

— Retour au texte —

6 – Note 24

Cf. J.-P. DUPUY, *Introduction aux sciences sociales. Logique des phénomènes collectifs,* Ellipses, Paris, 1992.

— Retour au texte —

6 – Note 25

Daniel DENNETT, *The Intentional Stance*, MIT Press, 1987 ; trad. fr. par Pascal Engel, *La Stratégie de l'interprète*, Gallimard, Paris, 1990.

— Retour au texte —

6 – Note 26

F. VARELA, E. THOMPSON et E. ROSCH, *L'Inscription corporelle de l'esprit*, Seuil, Paris, 1993.

— Retour au texte —

6 – Note 27

Traduction proposée par P. Engel.

— Retour au texte —

6 – Note 28

Daniel DENNETT, *Consciouness Explained*, Little, Brown and Compagny, 1991, p. 418 ; ma traduction ; trad. fr. par P. Engel, *La Conscience expliquée*, Odile Jacob, Paris, 1993.

— Retour au texte —

Index – Note

Afin de simplifier la lecture des recherches indexées, tapez ou copiez-collez dans l'outil recherche le nom de l'item choisi.

— Retour au texte —